IN MEMORIAM: DAN MARKEL

LOSING FRIENDS ................................................................. Garrick Pursley

A TRIBUTE TO PROFESSOR DAN MARKEL ................................. Keith L. Savino

DAN MARKEL’S PREMATURE DEATH CEMENTS 
His Uncompromising Legacy .................................................... Ryan Wechsler

ARTICLES

THE ILLUSION OF AUTONOMY IN
WOMEN’S MEDICAL DECISION-MAKING ...................................... Jamie R. Abrams

PARTITIONING AND RIGHTS:
The Supreme Court’s Accidental
JUSPRUDENCE OF DEMOCRATIC PROCESS ..................................... James A. Gardner

COPYRIGHT’S MERCANTILE TURN ............................................. Glynn S. Lunney, Jr

THE COST OF AVOIDANCE: PLURALISM, NEUTRALITY,
AND THE FOUNDATIONS OF MODERN LEGAL ETHICS ..................... Melissa Mortazavi

WORK MADE FOR HIRE—ANALYZING THE
MULTIFACTOR BALANCING TEST .................................................. Ryan Vacca

NOTES

BALANCING NATIONAL SECURITY POLICY:
Why Congress Must Assert its Constitutional
Check on Executive Power ...................................................... Rebecca Lightle

THE FAILURE AND FUTURE OF LAKE
OKEECHOBEE WATER RELEASES:
A QUASI-GOVERNMENTAL SOLUTION ........................................... Jacquelyn A. Thomas

VOLUME 42       FALL 2014       NUMBER 1
TABLE OF CONTENTS

IN MEMORIAM: DAN MARKEL

LOSING FRIENDS .............................................................. Garrick Pursley 1

A TRIBUTE TO PROFESSOR DAN MARKEL .................. Keith L. Savino 9

DAN MARKEL’S PREMATURE DEATH CEMENTS
HIS UNCOMPROMISING LEGACY ................................. Ryan Wechsler 13

ARTICLES

THE ILLUSION OF AUTONOMY IN WOMEN’S
MEDICAL DECISION-MAKING ................................. Jamie R. Abrams 17

PARTITIONING AND RIGHTS: THE SUPREME
COURT’S ACCIDENTAL JURISPRUDENCE
OF DEMOCRATIC PROCESS ................................. James A. Gardner 61

COPYRIGHT’S MERCANTILIST TURN .................. Glynn S. Lunney, Jr. 95

THE COST OF AVOIDANCE: PLURALISM,
NEutrality, AND THE FOUNDATIONS
OF MODERN LEGAL ETHICS ............................ Melissa Mortazavi 151

WORK MADE FOR HIRE—ANALYZING THE
MULTIFACtor BALANCING TEST .......................... Ryan Vacca 197

NOTES

BALANCING NATIONAL SECURITY POLICY: WHY
CONGRESS MUST ASSERT ITS CONSTITUTIONAL
CHECK ON EXECUTIVE POWER ............................ Rebecca Lightle 255

THE FAILURE AND FUTURE OF LAKE
OKEECHOBEE WATER RELEASES: A
QUASI-GOVERNMENTAL SOLUTION .................... Jacquelyn A. Thomas 285
Florida State University
College of Law

The Florida State University College of Law is fully approved by the American Bar Association and is a member of the Association of American Law Schools. The College offers a three-year course of study leading to the Juris Doctor degree with certificate programs in business, international and environmental law. Furthermore, a Master of Laws degree (LL.M.) in American Law for Foreign Lawyers is available for foreign graduate students who have legal training, and an LL.M. in Environmental Law and Policy is available. The law school also offers a one-year Juris Master degree program for those who possess a bachelor’s degree and who want to advance their careers with a year of legal studies and training.

The College of Law has approximately 700 students. Its research center has nearly 505,000 total print volumes and volume equivalents, 229,000 serials, and 3600 current subscriptions.

Educational opportunities are plentiful. The location of the College—only blocks from the Capitol, the Supreme Court of Florida, a number of federal and state courts, and various administrative agencies—makes extensive externship programs possible. Students publish the Law Review, the Journal of Land Use & Environmental Law, and the Journal of Transnational Law & Policy. Moot Court and Mock Trial teams have won national honors. Each summer, the College conducts a nationally recognized program at Oxford University, England.

These resources support the goal Dean Mason Ladd identified when the College was founded in 1966: “Its one aim is excellence in the preparation and education of students so that they will acquire the knowledge, comprehension, and judgment essential to the performance of legal services in a complex and rapidly changing society.”
FLORIDA STATE UNIVERSITY COLLEGE OF LAW

ADMINISTRATION

JOHN THRASHER, B.S., J.D., President of the University
SALLY E. McRORIE, B.A., M.A., Ph.D., Interim Provost and Executive Vice President for Academic Affairs
DONALD J. WEIDNER, B.S., J.D., Dean and Alumni Centennial Professor
MANUEL A. UTSET, JR., B.S., J.D., William & Catherine VanDercreek Professor and Associate Dean for Academic Affairs
MARK B. SEIDENFELD, B.A., M.A., J.D., Patricia A. Dore Professor of Administrative Law and Associate Dean for Research
DAVID L. MARKELL, B.A., J.D., Steven M. Goldstein Professor and Associate Dean for Environmental Programs
DAVID E. LANDAU, A.B., Ph.D., J.D., Mason Ladd Professor and Associate Dean for International Programs
NANCY L. BENAVIDES, B.S., J.D., Associate Dean for Student Affairs
ROSANNA CATALANO, B.S., J.D., Associate Dean for Placement & Director of the Business Law Certificate Program
JANIA DANAIS INGRAM, B.S., J.D., Assistant Dean for Student Affairs
CATHERINE J. MILLER, B.A., Assistant Dean for Administration
JEANNE B. CURTIN, B.S., M.B.A., J.D., Director of Development

FACULTY

FREDERICK M. ABBOTT, B.A., J.D., LL.M., Edward Ball Eminent Scholar
WENDI ADELSON, B.A., M.PHIL., J.D., Clinical Professor
KELLI A. ALCES, B.A., J.D., Loula Fuller and Dan Myers Professor
PAOLO ANNINO, B.A., M.A., J.D., Ph.D., Glass Professor of Public Interest Law
ROBERT E. ATKINSON, B.A., J.D., Greenspoon Marder Professor
SHAWN J. BAYERN, B.S., J.D., Larry and Joyce Beltz Professor of Torts
RICHARD M. BENHAM, B.S., J.D., Entrepreneur in Residence
BRUCE L. BENSON, B.A., M.A., Ph.D., Courtesy Professor
TAMARA G. BLENKORN, B.A., J.D., Legal Writing Professor
PATRICIA H. BORN, B.A., M.A., Ph.D., Courtesy Professor
COURTNEY CAHILL, B.A., Ph.D., J.D., Donald Hinkle Professor
DONNA R. CHRISTIE, B.S., J.D., Elizabeth C. & Clyde W. Atkinson Professor
TERENCE C. COONAN, B.A., J.D., Courtesy Professor
TALBOT D’ALEMBERTE, B.A., J.D., President Emeritus and Professor
SALLY C. GERTZ, B.A., J.D., Clinical Professor
ELWIN J. GRIFFITH, B.A., J.D., LL.M., Tallahassee Alumni Professor
MELISSA O. HANSON, B.S., J.D., Visiting Legal Writing Professor
SHI-LING HSU, B.S., M.S., Ph.D., J.D., Larson Professor
R. MARK ISAAC, B.S.F.S., M.S., Ph.D., Courtesy Professor
STEVE R. JOHNSON, B.A., J.D., University Professor
FAYE E. JONES, B.A., M.S., J.D., Director of the Research Center and Professor
JEFFREY H. KAHN, B.A., J.D., Charles W. Ehrhardt Professor
MARSHALL KAPP, B.A., M.P.H., J.D., Professor of Medicine & Law
JAY KESTEN, B.S., LL.B., LL.M., Assistant Professor
GARY KLECK, A.B., A.M., Ph.D., Courtesy Professor
LAWRENCE S. KRIEGER, A.B., J.D., Clinical Professor and Director of Clinical Externship Programs
DAVID E. LANDAU, A.B., Ph.D., J.D., Mason Ladd Professor and Associate Dean for International Programs
JENNIFER P. LAVIA, A.B., J.D., Legal Writing Professor
TAHIRIH V. LEE, A.B., A.M., A.M., M.PHIL., J.D., Ph.D., Associate Professor
JAKE LINFORD, B.A., J.D., Assistant Professor
DAVID LOGAN, B.A., M.A., J.D., Visiting Professor
WAYNE A. LOGAN, B.A., M.A., J.D., Gary & Sallyn Pajic Professor
BRUCE A. MARKELL, B.A., J.D., Jeffrey A. Stoops Professor
DAVID L. MARKELL, B.A., J.D., Steven M. Goldstein Professor and Associate Dean for Environmental Programs
TRICIA ANN MATTHEWS, B.A., M.A., J.D., Legal Writing Professor
GOLDBURN P. MAYNARD, JR., B.A., J.D., LL.M., Visiting Assistant Professor
MURAT C. MUNGAN, B.A., M.A., PH.D., J.D., Assistant Professor
SETH ORANBURG, B.A., J.D., Visiting Assistant Professor
GARRICK PURSLEY, B.A., J.D., Assistant Professor
JARROD F. REICH, B.A., J.D., Legal Writing Professor
JOHN SCHOLZ, B.A., M.S., PH.D., Courtesy Professor
MARK H. SEIDENFELD, B.A., M.A., J.D., Patricia A. Dore Professor of Administrative Law and Associate Dean for Research
MARK SPOTTWOOD, B.S., J.D., Assistant Professor
NAT S. STERN, A.B., J.D., John W. & Ashley E. Frost Professor
RUTH E. STONE, B.S., J.D., Wayne and Pat Hogan Professor of Trial Practice
CHARLIE MORGAN TAYLOR, B.A., J.D., Legal Writing Professor
FERNANDO R. TESON, J.D., LL.M., S.J.D., Tobias Simon Eminent Scholar
FRANITA TOLSON, B.A./B.S., J.D., Betty T. Ferguson Professor of Voting Rights
MANUEL A. UTSET, JR., B.S., J.D., William & Catherine VanDercreek Professor and Associate Dean for Academic Affairs
DONALD J. WEIDNER, B.S., J.D., Dean and Alumni Centennial Professor
ALLEN WINSOR, B.S., J.D., Richard W. Ervin Eminent Scholar Chair
HANNAH WISEMAN, A.B., J.D., Attorneys’ Title Professor
SAMUEL R. WISEMAN, B.A., J.D., McConnaughhay and Rissman Professor
MARY ZIEGLER, B.A., J.D., Stearns Weaver Miller Professor

RESEARCH CENTER STAFF
FAYE E. JONES, B.A., M.S., J.D., Director of Research Center and Professor
ELIZABETH FARRELL, B.S., J.D., M.S., Associate Director of Research Center
MARY M. MCCORMICK, B.A., M.L.I.S., J.D., Assistant Director for Research
MARGARET C. CLARK, B.A., M.A., Assistant Director for International Legal Research
KATIE MILLER, B.A., M.S., J.D., Assistant Director for Interlibrary Loan and Document Delivery
KATIE CRANDALL, B.A., M.S., J.D., Research and Outreach Librarian
AMY LIPFORD, B.A., M.S., J.D., Research and Access Librarian
BARBARA KAPLAN, A.B., M.S., J.D., Research and Scholarly Communications Librarian
PATRICIA BINGHAM-HARPER, B.A., B.M., M.S., M.M., Assistant Director for Cataloging
JON R. LUTZ, B.A., M.A., M.S., Assistant Director for Digital Projects

EMERITUS FACULTY
BARBARA ANN BANOFF, B.A., J.D., Professor Emeritus
JOSEPH M. DODGE, B.A., LL.B., LL.M., Professor Emeritus
CHARLES W. EHRRHARDT, B.S., J.D., Professor Emeritus
ROBERT H. KENNEDY, B.A., LL.B., Professor Emeritus
JOHN W. LARSON, A.B., J.D., Professor Emeritus
JARRETT C. OELTJEN, B.A., J.D., Professor Emeritus
DAVID F. POWELL, B.B.A., J.D., LL.M., Professor Emeritus
EDWIN M. SCHROEDER, PH.B., M.S., J.D., Professor Emeritus
HAROLD P. SOUTHERLAND, B.S., J.D., Professor Emeritus
JOHN W. VAN DOREN, A.B., LL.B., Professor Emeritus
JOHN K. VINSON, B.A., J.D., LL.M., Professor Emeritus
JOHN F. YETTER, B.S., J.D., LL.M., Professor Emeritus
The *Florida State University Law Review* is published four times per year at the College of Law, Tallahassee, FL 32306-1601. Periodicals postage paid at Tallahassee, Florida, and at additional mailing offices. United States Postal Service Publication Number 991780. ISSN 0096-3070. POSTMASTER: Send address corrections to Florida State University Law Review, College of Law, 425 W. Jefferson St., Tallahassee, FL 32306-1601.

The views expressed in published material are those of the authors and do not necessarily reflect the policies or opinions of the Law Review, its editors and staff, or the Florida State University. We welcome and will consider for publication letters to the Editor about work appearing in the Law Review.

Copyright © 2014 by the *Florida State University Law Review* Member, National Conference of Law Reviews

The *Florida State University Law Review* hereby grants permission for copies of all articles on which it holds the copyright to be made and used by nonprofit educational institutions, provided that copies are distributed at or below cost, the author and journal are identified, and proper notice is affixed to each copy. All other rights are reserved.

The *Florida State University Law Review* thanks Reed Elsevier for the donation of LEXIS research services.
General Information

Cite As. Cite this journal as FLA. ST. U. L. REV.

Submissions. The Florida State University Law Review welcomes unsolicited manuscripts on topics of interest to the legal profession. The Law Review acknowledges that unsolicited manuscripts may be submitted to other journals. All pieces will be evaluated as expeditiously as possible, but please submit an expedite request to the journal if necessary. Manuscripts should be typewritten in 12-point font, and footnotes are preferred. Citations should conform to The Bluebook: A Uniform System of Citation (19th ed.). Citations to Florida sources should conform to the Florida State University Law Review Florida Style Manual (7th ed.), a copy of which may be obtained for $15 from the Law Review. The Law Review accords great deference to an author’s work but reserves the right to determine the final form in which articles will be published.

Advertising. The Law Review accepts advertisements that may be of use to its readers. Rates for full- and half-page advertisements are available from the Editor-in-Chief at the mailing address and telephone number below.

Subscriptions. Annual subscriptions (four issues) are available for $30, and single issues are available for $10 each. Florida firms and individuals should add the appropriate sales tax. Foreign subscribers should add 50 cents per issue or $2 per year for postage. Subscriptions are renewed automatically unless the subscriber sends a timely notice of termination. All notifications of address changes should include old and new addresses with ZIP+4 codes. Article reprints are available for 15 cents per page by contacting the Office Manager. Back issues more than one year old may be purchased from William S. Hein & Co., 1285 Main Street, Buffalo, NY 14209 and are also available in electronic format at http://www.heinonline.org.

Correspondence. Address all correspondence to the Florida State University Law Review, College of Law, 425 W. Jefferson St., Tallahassee, FL 32306-1601. Direct manuscripts to “Article Submission.” Direct all other advertising and business matters to the Editor-in-Chief. Contact the editorial and business staff at (850) 644-2045.

Errata. Bring errors in this volume to the attention of the Editor-in-Chief. The Law Review will publish corrections when necessary.

Access. This document is available upon request in other formats for individuals with print-related disabilities. Contact the ADA Coordinator, FSU Office of Equal Opportunity and Compliance, 6200 University Center, Building A, Tallahassee, FL 32306-2410. Phone: (850) 645-1458 (Voice) or (850) 644-1546 (TDD). E-mail: eoc@admin.fsu.edu.

THE MISSION OF THE
FLORIDA STATE UNIVERSITY LAW REVIEW

To provide a forum for contemporary legal discourse, we must:

Adhere to the highest standards of analysis, accuracy, and timeliness in publishing the Law Review;

Distinguish the College of Law as a leading institution, known for its support of scholarship and academic pursuit;

Attract the College of Law’s finest legal communicators by encouraging independent legal thought, offering positions of responsibility, and maintaining a tradition of service to our College;

Inform our profession about the compelling issues of our time by providing a forum for student scholars, practitioners, and legal educators;

Offer our profession a selection of articles balanced between the theoretical and the practical;

Uphold the honor and decorum of our profession by championing the highest standards of conduct;

Stand together as advocates of quality legal education and strive to improve the reputation of our Law Review and of our school;

Inspire cooperation, commitment, and camaraderie among our membership and the College of Law at large; and

Safeguard an institutional memory and tradition so that those who follow will benefit from our experience.
LOSING FRIENDS

GARRICK PURSLEY*

I promise you this essay is about Dan Markel. It’s painful, still, to talk or to write about him. This is a memorial, a remembrance—so I’ll tell you how I remember Dan. I don’t know how, exactly, to explain his impact on my life. But it’s prototypical, I think, of the impact he had on the lives of many of the people who encountered him, so I’m going to try and back my way into it.

All my life I’ve wanted to be a professor. An academic, a scholar—someone who contributes to the world of ideas. To follow in the footsteps of my childhood idols—Bohr, Einstein, Newton, Descartes, Plato (I suppose that, of late, we should add Hart, Hand, and Posner to that list). Originally, I planned to earn a Ph.D. in philosophy. As a philosophy major at the University of Texas, it seemed the natural next step to find my way into the academic life. I had a romantic—probably overly rosy and idealized—vision of the academic life. I saw myself attending conferences and dinner parties with lively discussions of the work of thought; sitting in an office, lit warmly by table lamps and surrounded by books, chin in hand, contemplating and staring with portent into the foreground; wearing tweed jackets; writing something meaningful—a book, perhaps. I went to law school after a conversation with my father on the front porch of my college apartment. Beers in hand, Dad told me what his research had revealed about the job prospects for philosophy Ph.D.s. He suggested law school. I had no idea that I’d be any good at it. I didn’t want to

* Associate Professor, Florida State University College of Law. For whatever it’s worth, this is for Ben and Lincoln. I’m grateful to my FSU Law colleagues, and especially to Jeannie Wanzek, for holding me up in recent months.
spend a lifetime looking for work and doing whatever would pay the bills in the meantime. So I went to law school.

I’m still not sure why, but I applied to only a couple of law schools, one of which was Texas—it’s where my father went to medical school (he was a big Texas partisan); and of course, because I was (1) a current University of Texas student, (2) a Texas resident, and (3) a legacy, it made perfect practical sense. I thought I’d get in even though I’d only become serious about my work in the later years of college. I didn’t apply to many other places—Baylor accepted me; SMU, and a couple others. Tulane, as I recall. On a lark, I applied to Harvard Law School, but I made only a halfhearted effort on the essay and was waitlisted. Harvard, of course, is where Danny went to school. College, law school—he was the very model of a modern Harvard person (except for his years in Cambridge earning a master’s—Cambridge is where Newton held the Lucasian Professorship in Mathematics, which Stephen Hawking now holds). Danny was, by the time I was admitted to law school in 2000, already a Harvard legend—editor of the *The Crimson*, editor of the *Harvard Law Review*—doing things I didn’t remotely have the imagination to imagine. I suppose if I’d applied to Stanford or Columbia, I might’ve been accepted and ended up with a better résumé for the legal academic job market. Life, as you all know, is composed of thousands of contingencies; we can never know what might’ve happened if we’d decided this or that thing some other way. I started law school at Texas in 2001. At that point, I didn’t know how much of a problem it was that I hadn’t tried harder to get into Harvard or Yale. After the first semester, I discovered I was good at law school (there was some logic to it, and some philosophy in it) and proceeded to find an energy I’d never known. I talked to professors and discovered, to my delight, that they’d also gone to law school and, so, perhaps I might be able to enter academia, after all. Of course, if one doesn’t graduate from Harvard or Yale, it’s an exercise in ice skating uphill. But by God, I was going to do everything in my power to make it happen. So I did everything—law review, research assistantships, as many good grades as I could earn—all with my eyes toward becoming a law professor. Someday.

Those law school years involved a great deal of planning and projection—I always tried to make choices that would leave open all options that might be useful on the academic path. Clerking for a federal judge seemed like the best option—Danny, after all, had clerked right out of law school for Judge Mike Hawkins on the Ninth Circuit. So I took a job as a law clerk for Judge Royce Lamberth in Washington, D.C. I wrote my first real law review article that year. Writing it reminded me of my larger goal; and trying to get it published reminded me how hard it would be. That year, I began reading
PrawfsBlawg and discovered Dan; he was, of course, one of the progenitors of legal academic blogging. I sent him an e-mail that year, asking for advice on how to make a respectable run at legal academia. He must have received a couple hundred e-mails each week—everyone reading this knows that by 2005 (when I was clerking and pestering him) he had something like a billion friends and was already on the path I wanted to walk. He responded. I can’t find the e-mail—at that time, I was using some free e-mail service to which I no longer have access. Contingencies. I’d go back and print out that e-mail if I could. It would be several years before we communicated again. What I remember, though, is how patient he was—he took the time to think about my particular qualifications, my particular circumstances, my particular goals. And he advised me well: “Get an appellate clerkship and write something else.” I did as he advised, clerking for Judge Tim Dyk of the Federal Circuit. I corresponded briefly with Dan during that year. He said he was proud of me.

By the time I left D.C. in 2006, Dan had finished his first year as a professor at the Florida State University College of Law. He’d published seven law review articles, including in the Minnesota Law Review and the Vanderbilt Law Review; and he’d worked on several manuscripts that would, in 2007, be published in similarly prestigious journals. Additionally, he’d started laying the intellectual foundation for his book. I was intimidated, daunted, but determined to try and follow his lead.

After my clerkships, I joined a law firm in Dallas and published a second law review article. After a while, several people contacted me about the possibility of a fellowship on the University of Texas School of Law faculty (in what was then called the “Emerging Scholars Program”). I was excited and scared—it was my alma mater; and it was hard to think about calling my own professors by their first names. I asked Dan what he’d do; he said a fellowship could only help, given my non-traditional résumé. I was always worried about that damned, non-traditional résumé. I packed up and moved back to Austin. That was the summer of 2008. Dan said, as I recall his e-mail, “You’re on your way.” (I wish I had those e-mails). I was a regular Prawfs reader by this point, and there was something about that blog—there still is, I suppose; you could feel Dan in there, in between the lines, almost like he was in the room with you. Maybe it was in the way he wrote or the way he shared personal things. I’m not sure. But as I finally began the journey I’d prepared for and hoped for after so many years, I felt like Dan and I were already colleagues—perhaps even friends—though we’d never met. This was one of his special gifts—he could make you feel that way.

After the fellowship, I took my first full-time teaching position at the University of Toledo College of Law. Dan (by e-mail) was thrilled
that I’d landed anywhere at all (don’t forget my non-traditional résumé). Much of the credit goes to the wonderful people on the Texas Law faculty who supported me through the process. But Dan was there, too, sending occasional notes of encouragement and answering several relatively panicked e-mails during the weekend of the hiring conference (ironically, although I was back in D.C., I didn’t feel at all comfortable—they don’t call it the “meet market” for nothing). Dan was going up for tenure at this point, and of course, his was perhaps the easiest tenure case in the history of FSU. It was his style to remember details about people he barely knew—like me—and to find time to stay in touch. He was always positive; and his commitment to his relationships—even these quasi-relationships—amazed me. Toledo was nice—wonderful people and a very nice way to start an academic career. I wrote; I submitted articles to law reviews; I pestered Dan with questions about that process (especially about how to get good placements). Around this time, Prawfs began hosting an annual “angsting thread” for law professors subjecting themselves to the law review submission process. I read this religiously; though I’m not sure how much it helped (I am sure that it created an angsty echo-chamber effect). It helped with the angst, and it felt like a community. Dan brought us together, and he didn’t discriminate according to résumé, school, or placement record. He was an equal opportunity friend.

The academic life in Toledo turned out to be a bit different from what I’d envisioned. Toledo was colder than any place I’d ever lived, and while my colleagues were kind and friendly, the faculty was small and everyone was busy. I didn’t know whether it was the place or my internal vision of academic life that was slightly off, but I needed to make a change. Once again, Dan was there; he helped me catch the attention of FSU Law’s appointments committee. After what I’m certain was a significant and very Dan-like effort to advocate on my behalf—he would never talk about it with me, but I’m sure Dan spoke on my behalf in faculty meetings, colleagues’ offices, the hallways, and so forth while my FSU application was pending—I was hired as an assistant professor at FSU Law, whose faculty is well-known for its scholarly bent. I was thrilled. This was the kind of place I’d always envisioned. And much warmer, with beaches. I am only here because of Dan Markel. Had he not responded to my e-mail—and that’s really all it was, I just sent him an e-mail that said something like “do you think I’d have a chance of getting an interview at FSU?”—I have no idea what would have happened or where I would be. Dan, of course, always responded to e-mails; he always helped when he could. And even as I remain convinced that our life paths are radically contingent and determined by an enormous number of factors, any one of which, if tweaked, could dramatically alter
the outcome; I am also convinced that there’s no possible world in which Dan wouldn’t have done what he did for me. That’s the kind of thing he did for his friends. I’ll never have the chance to repay him.

That, then, is how I came to be Dan’s colleague. In the short two years in which I had that honor, I learned two important lessons from Dan that I want to share.

First, by the time I joined the FSU Law faculty, Dan was one of the more famous law professors in the country. There was his path-breaking blog, his brilliant work on criminal punishment, his appearance at what had to have been a record number of conferences, and his continuous contributions to the media. I worked constantly to try to emulate him: I shut myself up in my new FSU office and wrote, and wrote, and wrote. I came up for air to attend workshops and faculty meetings; and I worked hard to improve my teaching. I can’t say how many more hours I worked per week than at the law firm. To be tenured on a faculty like this and to eventually become prominent would require sacrifice, I rationalized. I didn’t go to my daughters’ preschool to have lunch with them or to pick them up early to go to the park—they stayed all the hours the preschool was open, and I stayed in my office each day until the very last minute past which I knew I couldn’t make it to the school by closing time. I still prided myself on being a good, involved father—my daughters and I did every kid-related thing in Tallahassee over the course of the first several months on weekend days.

But then I noticed that Dan handled things differently. Our kids went to the same school, and I saw him there from time to time, dropping off or picking up his sons. In fact, the first time I met Dan in person was when, on my house-hunting visit to Tallahassee after I’d accepted the faculty position, I went to that very school to meet the staff and to try and get my daughters admitted. He was there in the middle of the day, in his running gear after some hot summer jogging. He was standing in the parking lot when I pulled in; and it took me a minute to realize it was him. He was there to have lunch with his kids, which parents were always welcome to do at that school. The teachers told me that he came at least once per week and frequently more often than that, sharing spaghetti and meatballs, tacos, or whatever the kids were having; passing plastic dishes around; playing with them. He read stories to the kids; he visited on days when they were scheduled to learn about Jewish holidays and shared his faith with them. He wasn’t there just for his sons—he gave to all of the children in various ways; he was a regular fixture, almost like a part-time staff member. I realized that Dan had figured out some secret to balancing work and life; to being a professor and a father at the same time. I’m still not sure I’ve figured out that secret. He didn’t tell me what it was—he just modeled it for anyone who
cared to notice; that, I think, is one important lesson of Dan’s life. Dan was a role model for me in a number of ways: he taught me much about being a professor, a scholar, and a friend; but he taught me even more about how to be a whole person. I don’t quite have it down—my daughters have moved to a new school and parents aren’t allowed to come just anytime—but I’m still working on it, following his example. He was the best father I’ve ever known.

Second, Dan taught me about friendship and empathy. We both went through personal crises of sorts while we were on the faculty together—him first, then me. I won’t say much about the circumstances, but I will say this: in the midst of what I know was incredible pain and stress, Dan walked over to my office every few days, without my having to call for him, just to talk about it. He shared his experiences in the hope that they might help me. They did. I didn’t have a chance to tell him how much it meant—I was in a new place, going through something that rips people apart. There was a sense of falling, of the ground caving in from under my feet. Having someone who knew what it was like, and who was willing to listen and to share advice ranging from the emotional to the practical, made it possible for me to continue working, to make it through in one piece. The lesson is this: relationships matter; friendships matter—they matter more than publications, more than laurels and named professorships, and more than intellectual contributions or our impact on the larger world. Our lives gain their importance from the bonds we share with other people, I think. Dan took relationships seriously. He committed to them and nurtured them. He shared freely of himself, even the painful things. He had the gift of real empathy, and that was one of the ways in which he made a difference in the lives of others, including mine. Dan made enormous contributions to the world of ideas and to the community of legal scholars; he made a name for himself and made an impact on the world in a way that most people don’t. But what he taught me was intensely personal: pain is easier to bear when shared, and a little empathy goes a long way. I felt like I wasn’t alone. Dan taught me the lesson that now drives me: to make your life mean something, it needs to mean something to someone else. Of course, that sounds trite—we brush sentiments like this aside all of the time in the name of reason and truth—but it’s still true.

Life is a series of contingencies—we can’t know what would have happened had we decided this or that some other way. The crucial thing to remember is that whatever you currently like about your life is the result of an impossibly complex constellation of factors; each decision, in part, is a product of the last. To try and unravel them to say “this was, because I did that” or “had that not happened, I wouldn’t be here” is a fool’s errand. The accumulation of my deci-
sions—all of my planning and work—led me here. It led me to FSU; it led me to my fiancé; it led me to my work and to my calling. I cannot regret any of it, much as I might like to have had some things happen differently. Any tiny change could alter everything—like a pebble cascading down a mountain that eventually causes a landslide. If I could go back and save Danny, somehow, I would—he was the kind of which the world has too few. The world needs people like Dan. I miss him terribly, and I always will. I’ll ask myself whether I’m writing a problem paper or a puzzle paper; I’ll always play exuberantly with my daughters every chance I get; I’ll help people who are going through things that I’ve gone through; and I’ll always be grateful for the bits of luck that I’ve had here and there. Especially the luck of having known Dan, in his prime, while he was everyone’s friend. Even mine.
A TRIBUTE TO PROFESSOR DAN MARKEL

KEITH L. SAVINO*

The day I first met Professor Dan Markel I was prepared to walk into his office and tell him I was dropping his class. I was enrolled in his Sentencing Law and Policy seminar but had also recently been invited to become a member of the Florida State University Law Review. As a new member, I was required to write a lengthy student note, and I decided that completing papers for both Professor Markel’s seminar and the Law Review in the same semester would be too burdensome and well beyond my capacity. In my mind, dropping Professor Markel’s course was the obvious choice to lighten my workload.

After we exchanged introductions, I told Professor Markel about my dilemma and my “regretful” decision to drop his class. He told me that he understood my plight and was glad I came to speak with him. I was relieved that this awkward exchange was about to end, but Professor Markel continued. He looked at me and said: “The vast majority of law students in your position would do the exact same thing. They would take the easy road in the face of something difficult. Only a committed few would do both. Do you want to be in the majority, or do you want to see what you are capable of?” I was rendered speechless by Professor Markel’s proposal. He told me to take some more time to weigh my options and said that he would know my decision in two days on the first day of class.

In the time that has passed since Professor Markel’s shocking death, I have heard countless, similar stories of his unbridled passion and enthusiasm for pushing his students and colleagues to succeed and to reach their full potential. I decided to stay in Professor Markel’s Sentencing Law and Policy seminar, and it became one of my favorite, and most rewarding, courses in law school. I was among one of the final two classes that he would teach, and given the small, intimate nature of my five-person seminar, I developed a wonderful relationship with Professor Markel. His mentorship and advice throughout that semester transformed my writing and legal reasoning, and I owe so much of my growth and maturity as a future attorney to Professor Markel.

Experiencing a Dan Markel class lecture was truly witnessing a force of nature at work. Within a few weeks of the semester I was convinced that he was the most brilliant person I had ever met, and I was continually awed, and oftentimes completely confused, by the ease with which he could posit and discuss complex sentencing law theory. During one lecture in particular, Professor Markel intricately explained the facets of his retributivism sentencing theory as applied

* J.D. Candidate, 2015, Florida State University College of Law.
to the character Jean Valjean from *Les Misérables*. After an almost fifteen-minute-long monologue on how Valjean fit the retributivist principles, I asked Professor Markel if he often used Valjean as an example to explain retributivism. Professor Markel laughed and said that it was just something that came to him off the top of his head. Like I said—a force of nature at work.

Beyond the confines of the classroom, the truly remarkable Professor Markel revealed himself further. Throughout the semester, we would meet every Friday to talk about my paper and to hash out ideas on how to improve my writing. It was through these discussions, oftentimes sprawled out on the couches near his office, that I came to know Professor Markel as the man, legal scholar, and loving father that he was. We went through many stages of the writing process together, and I experienced the brutally honest critiques and comments that he is famous for. I did not know how to react when he told me that my first draft was “mediocre at best,” or when there was a giant red “X” through three pages; but we eventually graduated to, “now you’re showing me what you’re capable of,” and then, “you’ve created something very special and thought provoking.”

It was through Professor Markel’s relentless desire to push me and to demand my best that I truly expanded my ability to think critically and to challenge conventional norms through my writing. The lessons and techniques that I learned from working one-on-one with Professor Markel have transformed my legal scholarship, allowing me to reach beyond my self-imposed comfort zone, challenge myself, and emulate his work ethic. We would often joke that Professor Markel probably never slept, given the amount of things he could accomplish in one day; but it was obvious from working with him that he strived to be the best in his field and had a staunch commitment to his work that few can claim.

A tribute to Professor Markel would be incomplete without mentioning some anecdotes that capture his personality and zeal for life, and I would like to share a few.

While I was writing my sentencing law paper, it was not uncommon to get an e-mail from Professor Markel at midnight or 1:00 a.m. with links to law review articles or *New York Times* op-eds related to my paper topic. He would include a quick sentence saying: “Thinking about your paper, started looking up some materials and came across this. Hope it helps.” I was always struck by these e-mails, and they speak directly to Professor Markel’s investment in his students. He was unequivocally devoted to our growth and progress to the point where he would analyze and think of ways to improve our work, even at absurd hours of the night. It was instances like this that made me want to work even harder for Professor Markel and to put the same level of effort into my work.
Professor Markel is well known for connecting people and for sharing his vast network of friends and associates, and he spared me no exception. When he found out I would be interning in Washington, D.C. over the summer, he took it upon himself to contact his friends who lived in the city; send me a complete list of restaurants to try; and recommend the best places for live music, good drinks, and somewhere to unwind after work. I was working in Washington, D.C. when I found out that Professor Markel had been killed. I will never forget walking out to the park across from my apartment and sitting on a bench for hours—numb—not wanting to believe the news I had just received.

Finally, the largest and, most important aspect of Professor Markel’s life was his two sons and he used every opportunity to express his love and admiration for them. During one of our Friday sessions, Professor Markel was obviously distracted on his iPhone as he kept scrolling through something while wearing a huge smile on his face—one that expressed pure joy. Sensing my frustration, Professor Markel flipped his phone around and excitedly showed me pictures he had taken of a painting that Benjamin, his son, made. He told me that he was so overcome and happy by his son’s artwork that he could not stop looking at it. Professor Markel showed me more photos and boasted about how perfect his boys were—with that huge smile on his face the entire time. I am forever grateful to have witnessed this type of love a father can have for his sons.

The final time I saw Professor Markel, he invited our seminar class to his house for the traditional “final class dinner.” We were looking forward to this final class all semester, which included dinner, personally prepared by Professor Markel, and then a group peer review of our papers. When I got to his house, Professor Markel was bustling around the kitchen in his socks, putting a vegetable lasagna in the oven, making homemade guacamole, and filling wine glasses. It was impossible not to notice the string hanging across the living room ceiling, holding Benjamin’s and Lincoln’s vast array of artwork.

Given the unsettling tragedy surrounding Professor Markel’s death, I am comforted to have this positive, happy, and final memory of him. We spent the evening laughing over dinner and had a very thoughtful and encouraging peer review, led by Professor Markel’s uncanny ability to facilitate discussion. I shook Professor Markel’s hand as I left and he reminded me that if I ever needed anything over the summer, or if had any questions, law school related or not, he was always there.

Although six months have passed since his death, I still have to catch myself from walking up to his third floor office in B.K. Roberts Hall, wanting to tell him a piece of news or to ask for his advice. I want to update him on things in my life and to ask him whether he
agrees with my decisions about classes, job interviews, and plans for my future. While I have to remind myself that these things are no longer possible, I carry with me the things he said, the e-mails he sent, and the values of hard work and devotion he instilled in me. I will never forget Professor Markel, and I will do my best to keep his memory and spirit alive.
“You know, they say to never be late with a cup of coffee in your hand.” Two seconds prior, I stumbled through the door of Professor Dan Markel’s Criminal Procedure: Adjudication class, thermos in hand, one or two minutes late. The door to our classroom was noisy, and the audience was limited to about ten, so all eyes pounced on me as I abruptly entered the room. Under Markel’s scrutiny, I immediately blurted my excuse: “traffic.” Although my excuse was true, Markel’s immediate and witty rebuttal proved it to be illegitimate—I should have skipped making coffee.

Brilliantly assertive, brutally honest, and tenaciously disciplined, Markel wore these attributes like a potent cologne. At the Florida State University College of Law where he taught, Professor Markel’s name carried a mystique that excited some students, terrified others, but marveled all. Before even meeting him, you would have heard the stories of Markel calling on the inattentive student who surfed the internet, interrogating the unprepared student who lied about completing the reading assignment, and dismissing the disruptive student who, from the back row, too loudly discussed the weekend’s football game with a classmate. Effectively, Markel’s expectations were disclosed before he ever handed out a syllabus, but those students who evaded campus gossip and were unaware of Markel’s demands quickly found out—and hopefully not the hard way.

To be fair, Markel’s dominant classroom command befall his brightest and most diligent students as well. I, admittedly, was not his brightest, but because of my passionate interest in criminal law, I was diligent in both classes I took with Markel, and I frequently contributed to classroom discussion. Typically, the unwritten rule in law school is this: law students who voluntarily raise their hand to participate get a “pass,” of sorts. More specifically, law professors usually restrict the harshness of their logical attacks on students’ volunteered comments in order to encourage voluntary classroom participation. Some variation of this rule was followed by every professor I took in law school except one, Professor Markel, who adamantly abandoned it.

Moreover, because Markel’s curriculum and teaching style were largely theoretical, his classroom discussions were prevalent and combative. When I was really on my game, and on days when my intuition was particularly strong, I could withstand two or three verbal
exchanges before Markel would stump me. But it was never about “winning” or “losing” when you went up against Markel; in fact, you were never against Markel whatsoever. It was a mutual learning experience, and that is what made his teaching style so fun and effective. Whenever I contributed, Markel would carefully consider my input (a brief delay in his rebuttal meant I might be on to something) and then meticulously unravel it. He treated students’ insights like pieces of scholarship to which he was obligated to critically respond. No one has ever unpacked my thought process in the way that Markel was so instinctively able to do. When I reflect on the times I went back and forth with Markel, I realize that they were the greatest moments of learning that I ever experienced. What I learned goes far beyond the intricacies of criminal law; Markel taught me how to think critically at a level I never had before, and that is a lesson Markel gave me that I will carry for the rest of my life.

Still, there was much more to Dan Markel than the militant professor and scrupulous scholar, and I am fortunate to have known the human being beneath the renowned academic. The first of two dinners that I had at Markel’s house revealed the contrast of who Markel was at home in comparison to who he was in the classroom. Markel invited the students in his Sentencing Law and Policy seminar, in which I was enrolled, to his home for dinner. I remember my curiosity as to what his home would be like. I boyishly (and presumptively) imagined his living room hosting exquisite furniture, his shelf space packed with scholarship, and his walls showcasing fine art and framed accolades. I am embarrassed to admit that I expected more of an intimidating library than a welcoming “home,” but my prediction nevertheless proved remarkably inaccurate. In fact, Markel’s home highlighted the family man in Dan Markel. Beanbag chairs and other children’s furniture lined his living room, toys littered his floor, and drawings and colorings decorated his walls. And whichever shelves were not flooded with coloring books and children’s stories, pictures of Markel’s family—primarily his children—were proudly displayed. Indeed, the makings of Markel’s home broadcasted not only that Markel was a playful, loving, and devoted father, but that fatherhood was by far the most important aspect of his life.

Furthermore, through the two classes I took with Markel, along with our countless interactions outside of the classroom, Markel and I developed a rather close relationship during my final two years of law school. I am proud to write that Markel became more than my professor; Markel was my friend and mentor. A personal honor both then and now, Markel rendered a special interest in me as a student and offered tangible assistance with my aspiring career in criminal defense. The impact Markel made on me is illuminated through many recollections I have of our relationship, but one final anecdote...
best encapsulates who Dan Markel was to me. Alongside my co-author and while juggling a heavy course load, I worked tirelessly for three months researching, writing, and editing a lengthy legal comment that I hoped to publish. Of course, I proudly boasted of the endeavor to Markel, but I maintained serious doubt that the paper would actually materialize into a publishable work. Months after finally completing the paper and soliciting it for publication, a specialty journal e-mailed me an offer for publication. Reading the offer (which my co-author and I readily accepted) was, by far, the proudest moment I ever had as a student; and I remember the thought that instantaneously popped into my mind as I exhilaratingly leaped from my chair: “I have to tell Markel!”

Whenever I contemplate what Markel meant to me, I reflect on that moment and how I subsequently darted to campus that day to look for him. He was someone I immensely admired and who I constantly strived to impress. It is for these reasons that I made plans with Markel to keep in touch as my career developed. Tragically, those plans were prematurely taken from our relationship. But what has not been taken is my desire to impress Markel, because I still feel that precise motivation, intuitively and as strong as ever before. From that standpoint, I am but one example of how Markel’s influence on the world has not lost momentum. No, Dan Markel is not merely a memory—his imprint is here to stay.
THE ILLUSION OF AUTONOMY IN WOMEN'S MEDICAL DECISION-MAKING

JAMIE R. ABRAMS

ABSTRACT

This Article considers why there is not more conflict between women and their doctors in obstetric decision-making. While patients in every other medical context have complete autonomy to refuse treatment against medical advice, elect high-risk courses of action, and prioritize their own interests above any other decision-making metric, childbirth is viewed anomalously because of the duty to the fetus that the state and the doctor owe at birth. Many feminist scholars have analyzed the complex resolution of these conflicts when they arise, particularly when the state threatens to intervene to override the birthing woman's autonomy. This Article instead considers the far more common scenario when women and their doctors align in the face of great decision-making complexity and uncertainty. What decision-making framework normalizes this doctor-patient alignment, and how does this decision-making framework complicate the actualization of autonomy for the women who do not elect this framework? This Article concludes that many, if not most, of the four million women who birth in hospital settings attended by physicians align with their doctors by applying a shared decision-making framework that presumptively elects the outcome that minimizes any, even minor, risks to the fetus. While individual patients can certainly elect this approach autonomously, when understood in the context of tort law—in which the actions of "most women" and "most doctors" can become the standard of care itself—this framework is deeply concerning.

This fetal-focused decision-making framework perpetuates an illusion of autonomy because doctors can apply the framework independently and universally. This decision-making model problematically resurrects the ghost of Roe v. Wade’s medical model in which doctors effectuate decision-making autonomy for women. Understood through a tort law lens, while this illusion of autonomy might not seem problematic to the individual woman who elect this framework, it risks imputing a distorted standard of care to all obstetric cases by creating a primacy that always prioritizes fetal risks over maternal risks, a primacy that explicitly contravenes existing tort standards. Tort law ordinarily governs "unreasonable risks," whereas this framework elevates any fetal risk to an unreasonable risk and reduces any maternal risk short of death to reasonable. It risks imputing to all women a standard requiring the complete acceptance of medical guidance.

This Article concludes that tort law standards should explicitly govern not just the "what" of childbirth outcomes, but the "how" of childbirth decision-making by using decision-making aids to ensure that women's autonomy is actual and not illusory. Incorporating decision-making aids in the standard of care would remedy the illusion of autonomy by

* Jamie R. Abrams (LL.M., Columbia University School of Law; J.D., American University Washington College of Law; B.A., Indiana University—Bloomington) is an Assistant Professor of Law at the University of Louisville Louis D. Brandeis School of Law. The author extends her sincerest thanks to Beth Burkstrand-Reid, Tim Hall, Meredith Harbach, Anibal Rosario Lebron, Ariana Levinson, Kim Mutcherson, Elizabeth Pendo, Jason Pletcher, JoAnne Sweeney, and Jessica Waters for feedback on earlier drafts. The author thanks participants in the conference Beyond Roe: Reproductive Justice in a Changing World (Rutgers School of Law–Camden 2013); the University of Louisville Louis D. Brandeis School of Law Faculty Workshop Series (2013); the SEALS Junior Scholars Workshop (2013); the Emerging Family Law Scholars and Teachers Workshops (Brooklyn Law School 2013 and the University of Minnesota Law School 2014); the Law & Society Feminist Legal Theory Network participants (Boston 2013 and Minneapolis 2014); and the Midwest Political Science Association Conference (Chicago 2014) for feedback and discussion on earlier drafts. The author also thanks the Brandeis School of Law Library, Carol Allen, Hallye Bankson, Kristin Birkhold, Stephanie Carr, Kim Balkcom, Annie Malka, and Corey Shiffman for invaluable research and editing assistance.
ensuring that “most women’s” decision-making frameworks are not presumptively applied to all women so as to distort tort law and undermine patient autonomy.

I. INTRODUCTION ................................................................. 18

II. NORMALCY AND CONFORMITY DOMINATE CHILDBIRTH .... 21
   A. The Normalized Medical Interventionist Model .................... 21
   B. The Propensity for Frequent Conflict .................................. 27
   C. The Relative Normalcy of Alignment and Absence of Conflict .. 28

III. HOW “MOST WOMEN” MAKE DECISIONS IN CHILDBIRTH .... 31
   A. Existing Accounts of Decision-Making in Childbirth ............. 31
   B. The Shared Fetal-Focus Framework ..................................... 37

IV. WHY WHAT “MOST WOMEN” DO IS CRITICAL TO AUTONOMY FOR ALL WOMEN 40

V. HOW DECISION-MAKING MODELS TO MINIMIZE FETAL HARM CREATE AN ILLUSION OF AUTONOMY ............................................................. 43

VI. WHEN THE DECISION-MAKING FRAMEWORK OF MOST WOMEN IS APPLIED TO ALL WOMEN IT PROBLEMATICALLY ALTERS MEDICAL STANDARDS OF CARE .... 47
   A. Minimization of All Fetal Harms Instead of Unreasonable Risks .... 47
   B. Complete Compliance with Medical Advice Has Never Been a Standard of Care .................................. 48
   C. The Ghost of Roe v. Wade’s Medical Model .......................... 50

VII. RE-WRITING AND RE-RIGHTING OBSTETRIC STANDARDS OF CARE .... 53
   A. Methodological Standards of Care ....................................... 53
   B. Using Decision-Making Aids ............................................... 55

VIII. CONCLUSION ................................................................. 59

I. INTRODUCTION

This Article occupies an uncomfortable, but necessary, place for women’s rights—it considers how “most women” navigate medical decision-making in childbirth. It considers why there is not more conflict between women and their doctors over medical decision-making in childbirth. It does so to reveal the critical importance of the tort law lens to actualizing women’s birthing autonomy.

While patients in every other medical context have complete autonomy to refuse treatment against medical advice, elect high-risk courses of action, and prioritize their own interests above any other decision-making metric, childbirth is viewed anomalously because of the duty to the fetus that the state and the doctor owe at birth.1 These duties have led to excessive medical interventions,2 forced medical procedures, and criminal prosecutions against pregnant and birthing


women. Reproductive rights scholarship has defended women’s autonomy in medical decision-making when conflicts arise between women and their doctors, women and the state, and women and their fetuses. This scholarship lens examining conflict is vital to women’s autonomy.

This Article instead considers the far more common scenario when women and their doctors align in the face of great decision-making complexity and uncertainty. Despite historical advocacy for choice, most women regularly enter the most expensive and interventionist childbirth health care system in the world with great normalcy. Most medicalized hospital births still lack adequate informed con-


4. See generally Michelle Oberman, Mothers and Doctors’ Orders: Unmasking the Doctor’s Fiduciary Role in Maternal-Fetal Conflicts, 94 NW. U. L. REV. 451, 451 (2000) (explaining how conflicts between women and doctors can arise at any time from conception to birth on issues ranging from testing to delivery methods).


6. See Oberman, supra note 4, at 472 (explaining how the fetus is seen as a “second patient who faces greater risks of serious morbidity and mortality than does the mother”).

7. CARSON STRONG, ETHICS IN REPRODUCTIVE AND PERINATAL MEDICINE 2 (1997).

8. See WAGNER, supra note 2, at 9 (concluding that Americans pay “more per capita for maternity services than any other country in the world”); RICHARD W. WERTZ & DOROTHY C. WERTZ, LYING-IN: A HISTORY OF CHILDBIRTH IN AMERICA 63, 141 (1977) (explaining the historically interventionist role that doctors have played in childbirth); Elisabeth Rosenthal, American Way of Birth, Costliest in the World, N.Y. TIMES (June 30, 2013), http://www.nytimes.com/2013/07/01/health/american-way-of-birth-costliest-in-the-world.html?pagewanted=all&_r=0 (explaining the “sticker shock” American women experience where charges for childbirth have tripled since 1996).
sent and are largely uncontested by the four million women who give birth in this manner each year. From a tort law lens, the birthing experience of “most women” greatly influences the governing standard of care applied to all women and is therefore a critical unexamined site for study.

What decision-making framework normalizes the frequency of doctor-patient alignment in obstetric care, and how does this decision-making framework complicate the actualization of autonomy for the women who do not elect this framework? This Article concludes that many women align with their doctors by applying a decision-making framework that always seeks to reduce all risks to the fetus regardless of maternal risks or materiality.

This decision-making framework might actualize the autonomy of the women who elect this approach, but it perpetuates an illusion of autonomy that is problematic to the women who do not elect this framework. This illusion of autonomy resurrects the ghost of Roe v. Wade’s medical model, in which doctors effectuate decision-making autonomy for women. Understood in a tort lens, this illusion of autonomy risks imputing a distorted standard of care to all obstetric cases by creating a primacy that always prioritizes fetal risks over maternal risks regardless of likelihood or severity, a primacy that is explicitly inconsistent with existing tort standards. Tort law ordinarily governs “unreasonable risks,” whereas this framework elevates any fetal risk to an unreasonable risk and reduces any maternal risk short of death to reasonable. It risks imputing to all women a standard that their autonomy is presumptively exercised by the complete acceptance of medical guidance.

This Article concludes that tort law standards should explicitly govern not just the “what” of childbirth outcomes, but the “how” of childbirth decision-making by using decision-making aids to ensure

9. Numerous feminist scholars and historians have chronicled and analyzed the absence of meaningful informed consent in childbirth interventions. Proper informed consent requires presenting the birthing woman with alternatives and medically accurate and complete information, including risks to the birthing woman, not just those to the fetus. See, e.g., Ketler, supra note 5, at 1031 (explaining that “the doctrine of informed consent was founded upon the notion that adult persons have a fundamental right to bodily self-determination”).


11. See infra Part III.B.
12. See infra Parts IV–V.
13. See infra Part VI.C.
14. See infra Parts VI.A–B.
15. See infra Part VI.A.
16. See infra Part VI.B.
that women’s autonomy is actual and not illusory. Incorporating decision-making aids in the standard of care would remedy the illusion of autonomy by ensuring that “most women’s” decision-making frameworks are not presumptively applied to all women so as to distort tort law and undermine patient autonomy.

II. NORMALCY AND CONFORMITY DOMINATE CHILDBIRTH

Despite the propensity for conflict and difference that childbirth seems to present, childbirth today is pervasively medicalized, hospitalized, and intervention-oriented. A.

The Normalized Medical Interventionist Model

Despite choice in birth methods, modern childbirth remains heavily normalized around a medicalized and intervention-oriented model. A survey on Listening to Mothers II considered women’s experiences with hospital births and concluded that “labor is literally pushed by routine or common measures” upon healthy populations through labor induction, augmentation, and direction, and it is also “pulled by interventions such as vacuum extraction/forceps, cesarean section, pulling on the cord to hasten birth of the placenta, and separation of babies from mothers after birth.”

Modern childbirth is “almost always” in a hospital. It is the leading cause of hospitalization today. Only 33,043 babies are born at home for every four million births in hospitals, while 98.7% of all

17. See infra Part VI.
18. Adele E. Laslie, Ethical Issues in Childbirth, 7 J. MED. & PHIL. 179, 181 (1982) (noting that this normalcy has been criticized for “imposing one model of treatment and care on individuals in widely differing circumstances,” yet conformity persists).
19. Advocates have sought to de-medicalize childbirth, but this view has not prevailed pervasively. DEBORAH LUPTON, MEDICINE AS CULTURE: ILLNESS, DISEASE AND THE BODY 154 (3d ed. 2012).
22. Laslie, supra note 18, at 185.
23. Sakala & Corry, supra note 21, at 11 (stating that six of the fifteen most common hospital procedures are related to childbirth).
24. Joyce A. Martin et al., Births: Final Data for 2011, 62 NAT’L VITAL STAT. REP. 1, 51 (2013), available at http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_01.pdf. However, some data positively suggest that the number of home births is increasing. Id. at 10 (indicating that “[t]he number of births occurring at home (33,043) [in 2011] was the highest since reporting began for this item in 1989” and that the number of out-of-hospital births attended by Certified Nurse Midwives also rose by 6% from 28.6% in 2005 to 30.2% in 2011); see also BRODSKY, supra note 10, at 177 (noting that the incidents of fetal death are identical in hospital and home deliveries).
babies are born in hospitals. This is a marked historical shift within the last century.

Modern childbirth is routinely overseen by physicians. Obstetricians hold a virtual “monopoly . . . over the maternity care system.” Of modern hospital births, 86.1% are performed by doctors of medicine, 7.6% by nurse midwives, and 5.8% by doctors of osteopathy. The medicalization of childbirth has dramatically exaggerated the role of doctors in birthing care and entrenched it. This is a modern continuation of “heroic” medicine traditions whereby physicians supplanted midwives and treated pregnancy with increasingly interventionist measures.

The rate of cesarean section births in the United States is particularly normalized. The rate of cesarean sections rose every year from 1996 to 2009, including a single year increase of seven percent. Although the use of cesarean deliveries seems to have remained steady in recent years, the procedure accounted for 32.8% of all registered births in the United States during 2011. About one in three babies is delivered by cesarean section today compared to one in five babies in 1996. With the increased cesarean rate comes the increased risks

25. Martin et al., supra note 24, at 10.

26. In 1900, “less than 5% of women delivered in the hospital.” Wertz & Wertz, supra note 8, at 133. In 1940, fifty-five percent of births were in hospital settings, and by 1960, eighty-eight percent were in hospital settings. Judith Walzer Leavitt, Brought to Bed: Childbearing in America 1750 to 1950, at 171 (1986). Generally, the non-hospital births were in rural areas. Id.

27. See, e.g., Wagner, supra note 2, at 5 (stating that obstetricians oversee ninety percent of American births and maintain a “monopoly . . . over the maternity care system,” while comparatively midwives oversee seventy-five percent of births in industrialized Western countries like Australia, Netherlands, and Great Britain). “[H]aving an obstetrical surgeon manage a normal birth is like having a pediatric surgeon babysit a normal two-year-old.” Id. at 5.

28. Martin et al., supra note 24, at 10. Notably, the number of hospital births attended by certified nurse midwives in 2011 reflects a six percent increase since 2005. Id. This increase also applies to the number of births attended by certified nurse midwives outside of the hospital. Id.

29. Wertz & Wertz, supra note 8, at 141 (“By 1920 doctors believed that ‘normal’ deliveries . . . were so rare as to be virtually nonexistent.”).


31. Martin et al., supra note 24, at 10. In 2011, vaginal deliveries constituted 2,651,428 of the 3,953,590 registered births in the United States. Id. at 52.

32. Id. at 10.

of health complications to women. As major abdominal surgery, cesarean births increase risks of infection and recovery complications to women.\textsuperscript{34}

Modern childbirth is almost universally reliant on medical interventions.\textsuperscript{35} Childbirth was historically a "natural" event.\textsuperscript{36} Doctors transformed childbirth over time into a series of "more precise and effective manipulations and interventions, both to prevent and to cure disease" which ensured that doctors were "on the lookout for trouble in birth."\textsuperscript{37} Modern birth is viewed as "something that cannot be left alone, that must be interfered with, monitored and 'helped along.'"\textsuperscript{38}

Modern birth is heavily managed in its timing and pacing. It is characterized by the frequent artificial rupturing of the water, induction and augmentation of labor, and managed pain treatment.\textsuperscript{39} Labor induction is the "use of drugs and/or techniques to cause labor to start, as opposed to waiting for labor to begin on its own through a complex interplay of maternal and fetal factors."\textsuperscript{40} The percentage of medically induced labors rose by 135% from 9.5% to 22.3% between 1990 and 2005.\textsuperscript{41} This has in turn contributed to earlier gestational
births for singleton babies from an average of forty gestational weeks to thirty-nine gestational weeks.\footnote{42}

Some women are requesting cesareans and inductions from their doctors independently.\footnote{43} Many others are undergoing these interventions without adequate informed consent at their doctor’s direction.\footnote{44} In the Listening to Mothers II survey of childbirth in U.S. hospitals, forty-one percent of women said that a health professional suggested inducing labor in eighty-four percent of the cases, and a total of thirty-four percent of the respondents actually had a medically induced labor.\footnote{45} Eleven percent of the respondents “felt pressure” to induce.\footnote{46}

Fetal monitoring technology has become standardized too. It is the most common obstetrical procedure performed in the United States.\footnote{47} While fetal monitoring technology emerged and garnered acceptance in the 1960s,\footnote{48} it was originally used only for high-risk pregnancies.\footnote{49} Today, electronic fetal monitoring technology “is the standard of care in virtually every community,”\footnote{50} despite persistent questions regarding its reliability and concerns regarding its basis for medicalized interventions.\footnote{51} Approximately eighty-five percent of all annual births in the United States use electronic fetal monitoring.\footnote{52}

\footnote{42. Id.}
\footnote{43. See Chris McCourt et al., Elective Cesarean Section and Decision Making: A Critical Review of the Literature, 34 BIRTH 65, 65 (2007) (identifying convenience, patient choice, and psychological factors, especially concerning negative experiences in prior childbirths and fear relating to childbirth, the perceived safety of a cesarean, and social and cultural factors).}
\footnote{44. “It is dubious that women have been sufficiently informed about the possible risks associated with artificial stimulation of labor, including over-stimulating the uterus, fetal distress, more painful contractions, and the cascade of procedures that may follow.” BRODSKY, supra note 10, at 143.}
\footnote{45. Sakala & Corry, supra note 23, at 36.}
\footnote{46. Id. at 37, 44 (listing consequences and side effects associated with inducing labor).}
\footnote{47. Thomas P. Sartwelle, Electronic Fetal Monitoring: A Bridge Too Far, 33 J. LEGAL MED. 313, 313 (2012).}
\footnote{48. Id.}
\footnote{49. ANGELA DAVIS, MODERN MOTHERHOOD: WOMEN AND FAMILY IN ENGLAND, C. 1945-2000, at 85 (Lynn Abrams et al. eds., 2012).}
\footnote{50. Sartwelle, supra note 47, at 313.}
\footnote{51. SHEILA KITZINGER, THE POLITICS OF BIRTH 46, 91 (2005); Sartwelle, supra note 47, at 313-14 (“[I]ts scientific foundation is feeble; inter-observer/intra-observer reliability is poor; the false-positive prediction of fetal distress rate is greater than ninety-nine percent; it has substantially increased the cesarean section rate with attendant mortality and morbidity; and it failed completely in its initial stated promise—reducing by half the incidence of cerebral palsy (CP), mental retardation (MR), and perinatal mortality.”).}
\footnote{52. Sartwelle, supra note 47, at 313; see also Sakala & Corry, supra note 21, at 27 (noting that women’s own accounts reveal a seventy-one percent usage of “continuous” fetal monitoring and another sixteen percent usage “most” of the time).}
And most of these interventions occur without proper informed consent.\textsuperscript{53} Doctors readily acknowledge informed consent models governing childbirth and pregnancy are different because of the treatment of the woman and the fetus.\textsuperscript{54} “Most of the time, medical interventions are employed without considering the woman’s choice or obtaining informed consent”; women merely sign a general permission of care form upon admission.\textsuperscript{55} Many women assume that the frequency and regularity of these interventions means that they are always in their best interest, unaware that “they may be exposed to avoidable and potentially harmful interventions . . . because of a lack of transparent comparative performance data to guide decisions and limited access to some effective high-value alternatives.”\textsuperscript{56} Women report wishing they knew more about the risks and side effects of these procedures.\textsuperscript{57} The pace of labor can complicate informed consent. Women report feeling dependent in labor on health professionals to make effective decisions “about which tests or procedures were in fact intrusive.”\textsuperscript{58}

And these interventions are heavily interconnected. “As one intervention justifies or increases the likelihood of using others, the cumulative effect is to create a distorted understanding of childbirth as a time when things are likely to go wrong and intensive medical management is required.”\textsuperscript{59} These interventions disrupt the natural process of birth and “incur a cascade of secondary interventions used to monitor, prevent, and treat the side effects of the initial interventions.”\textsuperscript{60}

These interventions are normalized and costly even though they may not conform to the standard of care. The medical costs of child-

\textsuperscript{53} See, e.g., BRODSKY, supra note 10, at 142-43 (contesting the “informed consent model” governing the doctor-patient relationship in childbirth); Ketler, supra note 5, at 1033 (explaining how underlying presumptions in historic cases position birthing women as “incompetent, irrational, ruled by nature, and therefore unable to make informed decisions” and noting that even modern cases position women as vulnerable and weak); Marjorie Maguire Shultz, From Informed Consent to Patient Choice: A New Protected Interest, 95 YALE L.J. 219, 221 (1985).

\textsuperscript{54} Oberman, supra note 4, at 472. See generally Pamela Harris, Compelled Medical Treatment of Pregnant Women: The Balancing of Maternal and Fetal Rights, 49 CLEV. ST. L. REV. 133, 134 (2001) (noting how informed consent is believed to be more complicated in childbirth because of the fetus).

\textsuperscript{55} BRODSKY, supra note 10, at 166; see also KITZINGER, supra note 51, at 46, 91; Andrew Iverson Almand, Note, A Mother’s Worst Nightmare, What’s Left Unsaid: The Lack of Informed Consent in Obstetrical Practices, 18 WM. & MARY J. WOMEN & L. 565, 593 (2012) (“Why are such seemingly material risks of drugs and procedures unconscionably being withheld from expectant mothers by obstetricians? Without a doubt, every mother would expect to be told the preceding information, yet so few actually receive it.”).

\textsuperscript{56} Angood et al., supra note 35, at S25.

\textsuperscript{57} Sakala & Corry, supra note 21, at 66.

\textsuperscript{58} Lazarus, supra note 36, at 37.

\textsuperscript{59} Sakala & Corry, supra note 21, at 28.

\textsuperscript{60} Id.
birth have risen by $3 billion annually from 1996 to 2009. Costs are particularly high for interventionist childbirths in hospital settings. Yet, critically, these interventions are not achieving better outcomes. The United States spends far more on medicalized childbirth, yet lags behind many countries in key indicators. The U.S. Department of Health and Human Services’ Healthy People 2010 report found that the United States is moving away from healthy birth weight targets and experiencing a rise in maternal mortality rates. This is particularly so for women of color and lower-income women. And women are having more physical and mental problems immediately after birth.

“Most women” experience a medicalized birth in a hospital setting with sub-standard informed consent. To the extent that maternal-doctor alignments are normalized within certain patterns, this raises the question of which cultural norms are being endorsed and sustained. Lisa Ikemoto explained this powerfully as the “Code of Perfect Pregnancy,” where essentialism prevails and acts to
direct the power of the state at women along race, class, and culture lines in the name of “protecting fetal interests.” The resulting narrow standard . . . has an effect beyond that of taking from women the authority to construct pregnancy and motherhood for themselves; it also eliminates the possibility of difference.

61. Vedantam, supra note 33 (noting how obstetricians may be paid more for cesarean sections).
62. Sakala & Corry, supra note 21, at 15, 47 (explaining that the average charge in 2005 ranged from $7000 for uncomplicated vaginal deliveries to $16,000 for complicated cesarean deliveries and that non-hospital births averaged $1624).
63. Angood et al., supra note 35, at S24 (“The United States spends far more than all other countries on health care, yet lags behind many on currently available global maternal and newborn indicators.”). Maternal and newborn hospital charges totaled $86 billion in 2006, far exceeding those of any other hospital condition. Id.
64. Sakala & Corry, supra note 21, at 3. The World Health Organization reports that twenty-nine nations have better rates for maternal mortality in childbirth, thirty-five nations have better rates for neonatal mortality, and twenty-three nations have lower rates of low birth weight births than the United States. Id. at 17 (reporting 2005 data for mortality rates and 2003 data for low birth weights).
65. See, e.g., Angood et al., supra note 35, at S27 (noting particularly that black, non-Hispanic women were increasing in negative health statistics for neonatal deaths, low birth weight infants, and other negative birthing outcomes).
66. Sakala & Corry, supra note 21, at 16 (indicating that women birthing in hospitals in 2005 reported high rates of new-onset physical and mental problems in the first two months after birth, with many problems persisting to six months or more postpartum).
67. See Ehrenreich & English, supra note 30, at 28-30 (documenting the class distinctions of pregnancy).
The relative normalcy of a dominant birthing experience therefore bears further examination.

B. The Propensity for Frequent Conflict

The uniqueness of obstetric care would seem to present the opportunity for more conflict. It is one of the only medical contexts in which a doctor considers possible liability to two potential litigants—the fetus and the birthing woman—and in which the state has expressed a clear interest and willingness to intervene.

Typical obstetric care involves a series of decisions made with imperfect information surrounding the simultaneous health risks facing both the fetus and the birthing woman. Obstetric decisions frequently involve medical considerations that threaten or invoke both the health of the pregnant woman and the fetus. These decisions include cesarean delivery or vaginal labor, electronic fetal monitoring, responses to breech positions, vaginal births after cesarean sections, and choice of pain management.

Obstetric medicine is an imperfect, judgment-based practice that responds to uncertainty. It relies heavily on science and skill, but at bottom, doctors are acting with informed judgment in resolving conflicts involving some degree of uncertainty. There is little consensus within the medical community regarding which services are essential to maternal care and which interventions actually improve health outcomes, which should cause more variation in medical decision-making. Even where there is medical consensus, that consensus does not necessarily match the realities of the medical care that is

69. See MANUAL OF OBSTETRICS 404 (Arthur T. Evans & Kenneth R. Niswander eds., 6th ed. 2000) (explaining that the “[m]anagement of labor should achieve delivery in a reasonable period of time while providing maternal support and avoiding any significant compromise to the mother or fetus”).

70. See generally Harris, supra note 54, at 158 (explaining that “[a]s this idea of maternal tort liability grows, a pregnant woman’s choices diminish and the state begins to play a role in her pregnancy”).

71. See Oberman, supra note 4, at 451 (explaining how conflicts between women and doctors can arise at any time from conception to birth on issues ranging from testing to delivery methods).


73. Law, supra note 20, at 366.

74. See KATHRYN MONTGOMERY, HOW DOCTORS THINK 3 (2006) (explaining how doctors draw on skill as well as judgment in making decisions).

75. Id. (noting how medical education teaches what is “known” in medicine and then the clinical apprenticeship prepares doctors to act in response to the uncertainty).

76. Angood et al., supra note 35, at S34 (documenting a “lack of consensus on a comprehensive package of essential maternity services that have been shown to improve health outcomes, and should be covered by public and private insurance,” which leads to “unwarranted variation in maternity care”).
provided. Some “[p]ractices that are disproved or appropriate for mothers and babies in limited circumstances are in wide use, and beneficial practices are underused.”

This variation in best practices should lead to more variety of patient choice. There are choices available, women are competent to make the choices, and “reasonable professionals, and hence reasonable patients, disagree” about which options are best. Literature available to birthing women describing the range of childbirth and child rearing perspectives also offer competing, even contradictory, theories that would suggest more disagreement or conflict to be resolved in doctor-patient relationships. Given the individuality of birth, the range of options and choices presented, and the lack of medical consensus in standards of care, why then is there not more conflict or disagreement between doctors and birthing women?

C. The Relative Normalcy of Alignment and Absence of Conflict

Yet, but for a few iconic cases, very few women actually sue or explicitly challenge this medicalized, interventionist model of childbirth. “Most women” do not explicitly object to these interventions contemporaneously or retroactively. Rather, women’s accounts of hospitalized childbirth and medicalized childbirth “indicate[] that they had a rather ambivalent response to” the hospitalization and medicalization itself. While “some women are alienated by their experience of medicali[z]ed birth,” many women across social classes

77. Sakala & Corry, supra note 21, at 1. “Many maternity practices that were originally developed to address specific problems have come to be used liberally and even routinely in healthy women”; these practices include labor induction, epidural analgesia, and cesarean sections. Id. at 4. “Available systematic reviews also do not support the routine use of other common maternity practices, including numerous prenatal tests and treatments, continuous electronic fetal monitoring, rupturing membranes during labor, and episiotomy.” Id.

78. Law, supra note 20, at 366.

79. See Davis, supra note 49, at 114 (noting the “tensions, ambiguities, and indeed the contradictions that are present in the women’s accounts” of caring for children).


81. See Jamie R. Abrams, Distorted and Diminished Tort Claims for Women, 34 Cardozo L. Rev. 1955 (2013) (concluding that women rarely sue for birthing harms); see also David M. Engel, Perception and Decision at the Threshold of Tort Law: Explaining the Infrequency of Claims, 62 DePaul L. Rev. 293, 293-94 (concluding that fewer than one in fourteen personal injury victims consults a lawyer, only one in fifty sues, and nine out of ten never contact injurer or insurance company of injurer). Sparse tort literature considers why this is so, although speculation considers money, time, and aggravation as possible explanations. Id. at 294.

82. Davis, supra note 49, at 107.
welcome medical intervention, if not management, and are quite satisfied with hospital deliveries. Women’s interviews describing epidural anesthesia and caesarean sections, for example, are not described as “turning points” in labor, but rather “just another procedure undergone.”

In the relatively rare cases when doctors and birthing women conflict in decision-making, the results of judicial intervention have been notably mixed and inconsistent. Where conflicts do arise, occasionally courts resolve the dispute between the birthing woman’s selected course of action and the doctor’s recommended course of action. Some courts have held that women’s decision-making autonomy is absolute, while others have said that the rights of the fetus or the state override her rights. This uncertainty—particularly when it derives from high-profile cases—emphasizes a divide, which leaves women with little clarity regarding their birthing rights. The law certainly recognizes a strong presumption in favor of maternal autonomy, but that autonomy is far from absolute.


[W]omen “regardless of social class or ethnicity . . . spoke about childbirth as a natural process, but at least to some degree, they accepted the medical view of birth: that any number of things could go wrong and that ultimately they had to rely on authoritative knowledge and concomitant technological expertise of their physician to ensure that they had done everything possible to have a healthy baby.”

Id. at 73 (quoting Ellen Lazarus, What Do Women Want? Issues of Choice, Control, and Class in American Pregnancy and Childbirth, in CHILDBIRTH AND AUTHORITATIVE KNOWLEDGE: CROSS-CULTURAL PERSPECTIVES 132, 133 (Robbie E. Davis-Floyd & Carolyn F. Sargent eds., 1997)). That is certainly not to say that women are universally happy with their birthing experiences. Women do state that they want more information and they are not always happy with the relationships underlying the care they are given, a point that might ordinarily increase the likelihood for suit. Angela Davis’s account in Modern Motherhood explains that women surveyed about their birthing experiences did “often report[] their unhappiness with obstetric interventions,” but she concludes that this was “as much a criticism of the lack of information they received, the lack of choice they felt that they had in their care, and their dissatisfaction with their medical attendants (doctors, midwives, and nurses), as their dislike of the procedures themselves.” DAVIS, supra note 49, at 107. Women might report that they received “excellent medical care” overall but still criticize the interpersonal treatment they received from hospital staff. Id.

84. DAVIS, supra note 49, at 98.

85. See Hoffman & Miller, supra note 1, at 280 (concluding that “[d]ifferent state courts have issued many competing decisions, which emphasizes a lack of unification in this area of law”).

86. Oberman, supra note 4, at 451.

87. Hoffman & Miller, supra note 1, at 288.

88. Id. at 289.

89. See, e.g., STRONG, supra note 7, at 183. The author concludes as follows:

The bodily integrity of mentally competent individuals who are persons in the descriptive sense is an extremely important ethical value. Control over one’s
itself creates "confusion for women concerning the scope of their legal protections."  

Constitutional approaches addressing women’s decision-making autonomy and state interventions do not translate effectively into the private clinical setting of obstetric medical care. Nor have constitutional frameworks yielded consistent outcomes. Even cases that are deeply enshrined as beacons of patient autonomy, such as Schreiber v. Physicians Insurance Co. of Wisconsin, do not provide workable guidelines explaining when doctors should follow maternal decision-making and when they can override it.

This inconsistency and lack of clarity, however, has not necessitated or yielded any explicit tort standard or medical standard to address the anomalous nature of childbirth. At the moment of birth, doctors owe a duty of care to both the birthing woman and the fetus, and the doctor can be sued by either. Nowhere in tort literature or precedent is the complexity of childbirth decision-making fleshed out in a primacy lens clarifying how doctors should respond if these duties conflict. Nowhere in tort literature or precedent is a workable methodology presented for resolving disputes that might arise from decision-making conflict between the woman’s autonomy and her doctor’s duty to the fetus in birth, revealing the relative normalcy of alignments and rarity of conflict.

Likewise obstetric training texts generally make no mention whatsoever of the possibility of conflict or resolution of it, further supporting the normalcy of alignment and rarity of conflict. For example, Williams’ Obstetrics text, a leading text in obstetric practice,
makes no mention of conflicting duties or potential ethical conflicts that may arise between women and their doctors. 96 Foundations of Maternal & Pediatric Nursing, a foundational text used to train students in maternal and pediatric nursing, says nothing whatsoever about the distinction or complexities of managing two duties at birth and the conflicts this might present. 97 In Charles R.B. Beckmann et al.'s Obstetrics and Gynecology, there is a general statement that the rights of the woman and the fetus create ethical considerations that a doctor must resolve, but no guidance about how to resolve issues. 98 And this is not occurring regularly in clinical instruction either. Rather, “little time is spent in resident programs in medical specialties on bioethics, informed consent, professional responsibility, and communicating with patients.” 99

Modern birth is extremely normalized toward the alignment of women and their doctors. The question then becomes, what decision-making framework are women and doctors applying to yield this alignment?

III. HOW “MOST WOMEN” MAKE DECISIONS IN CHILDBIRTH

This section considers possible explanations for the normalcy of maternal-doctor alignments in obstetric decision-making. It first highlights briefly the existing accounts of women’s decision-making in childbirth. It then concludes that a shared framework in which women and doctors align to focus on the minimization of all fetal risks accounts for much of women’s alignment with their doctors. The remainder of this Article will consider the implications of this fetal-focus.

A. Existing Accounts of Decision-Making in Childbirth

One explanation for the normalcy of women-doctor alignments is that it reflects the ongoing subordination of women. Theories reflecting the subordination of women in reproduction are well documented in feminist scholarship and women’s history. This explanation defines women’s subordination by their reproductive function. 100 Innumerable historical examples exist of childbirth as subordination, par-
particularly as birth moved into hospitals and during the “Twilight Sleep” movement.\textsuperscript{101}

The reproductive subordination of women has particularly targeted women of color and poor women.\textsuperscript{102} Powerful historical accounts exist of doctors forcing and coercing sterilizations on poor women.\textsuperscript{103} Lynn Paltrow’s modern pioneering work documents the race and class distinctions of forced interventions today. She concludes that “low-income women and women of color, especially African American women, are overrepresented among those who have been arrested or subjected to equivalent deprivations of liberty.”\textsuperscript{104} Fifty-nine percent of the forced interventions were on women of color and seventy-one percent were on economically disadvantaged women.\textsuperscript{105}

Modern reproduction subordination is less about women’s subordination to doctors and more about women’s subordination to their fetuses.\textsuperscript{106} The fetus is “the newest ‘social actor’ in the American conservative imagination.”\textsuperscript{107} Some modern political framings have posi-

---

\textsuperscript{101} Hospital births and professionalized medicine distinctly shifted the balance of power, pushing out women’s domestic support systems. See LEAVITT, supra note 26, at 181, 190. In the mid-nineteenth century obstetric care model, women were “willingly submitting their bodies to their physicians without questioning,” and “[i]nstead of women birthing their babies, their babies were ‘delivered’ from them.” BRODSKY, supra note 10, at 7-8. In the ‘Twilight Sleep’ movement characterized by the heavy use of sedatives, women were “knocked out while their babies were ‘dragged out’ ” by obstetricians, and their babies were born “floppy,” “sedated,” and difficult to be stimulated. Id.

\textsuperscript{102} See generally SIMONE M. CARON, WHO CHOOSES? (2008) (chronicling the troubling history of racialized interventions in the reproductive choices of the poor and of African-American women); DANIELS, supra note 72, at 53 (concluding that women of color or in lower economic status are “more likely to be subject to forced medical treatments”); DOROTHY ROBERTS, FATAL INVENTION: HOW SCIENCE, POLITICS, AND BIG BUSINESS RE-CREATE RACE IN THE TWENTY-FIRST CENTURY 102 (2011) (chronicling how medical stereotyping leads to unequal access to high quality medical care and concluding that “[b]lacks are less likely to get desirable medical interventions and more likely to get undesirable interventions that good medical care would avoid”); Robin Fretwell Wilson, Autonomy Suspended: Using Female Patients to Teach Intimate Exams Without Their Knowledge or Consent, 8 J. HEALTH CARE L. & POL’Y, 240, 263 (2005) (describing the performance of intimate exams on patients absent full consent as a phenomenon that has "short-circuited the ethical sensitivity of many medical educators, who clutch to a variety of rationales for dispensing with the simple step of disclosing forthrightly the educational nature of practice procedures and asking permission").

\textsuperscript{103} CARON, supra note 102, at 213-14 (describing a high-profile account of Carol Brown, a woman in South Carolina who was pregnant with her fifth child and could not find a doctor in her town to deliver her baby unless she agreed to forced sterilization).

\textsuperscript{104} Paltrow & Flavin, supra note 3, at 300-01 (examining more than four hundred cases of arrests, detentions and forced interventions in forty-four states from 1973 to 2005).

\textsuperscript{105} Id. at 311.

\textsuperscript{106} See DANIELS, supra note 72, at 49 (explaining how physicians and hospital administrators have become “much more inclined to compromise the patient’s right to autonomy in the interests of fetal health,” and they “lean heavily in favor of forced medical treatment” under the guise of “saving [fetal] life”).

\textsuperscript{107} Id. at 3, 9 (explaining that this political framing emerged in the 1980s from a convergence of “cultural, political, legal, and technological developments,” which collectively brought “the fetus into the public consciousness as an independent and autonomous being”).
tioned the fetus as “‘housed’ inside the pregnant woman’s body” in which it can become “victimized by the woman’s neglect, ignorance, or abuse.”\(^\text{108}\) This fetal characterization “‘reduces women to incubators’ who are seen not as ‘full-fledged human beings, but merely better or worse vessels for fetuses.’”\(^\text{109}\) Certainly subordination still exists, yet this account cannot fully explain the normalcy of alignment in decision-making because it fails to account for increased women’s autonomy, and it needs to contemplate changing political and social conceptions of the fetus.

On the other extreme, is some measure of the normalcy of women-doctor alignments explained by the success of the women’s movement and consumer health movement securing women’s decision-making autonomy? Doctor-patient relationships were historically more paternalistic, particularly in childbirth.\(^\text{110}\) Activists successfully challenged this model of care in the 1970s and strengthened women’s active decision-making through informed consent.\(^\text{111}\) The “authoritarian physicians” of times past are being replaced by “doctors who enthusiastically support, or at least accept, the self-motivated patients who seek out information for themselves.”\(^\text{112}\) Indeed, “[r]espect for auton-

\(^{108}\) Id. at 28 (explaining how conservative politics have depicted the fetus as the “victim” of women’s “excesses and freedoms”); UPTON, supra note 19, at 166 (“The pregnant woman is increasingly portrayed as separate to and the adversary of her own pregnancy/fetus, by presenting a ‘hostile’ maternal environment or refusing proposed medical intervention.”).


\(^{110}\) See MARTIN L. PERNOLL, BENSON AND PERNOLL’S HANDBOOK OF OBSTETRICS & GYNECOLOGY 1 (10th ed. 2001) (noting that “the paternalistic care model,” which gave the physician the right to determine how much information a patient received about her condition and possible treatments, is “waning”); Law, supra note 20, at 363-64 (describing how “[t]raditions of paternalism and disrespect for patient choice” permeated the childbirth experience as women’s care became routine in hospitalized settings by the 1950s, historically including sedation, removal by forceps, episiotomies to facilitate the forceps, and restraints).

\(^{111}\) See generally Holly Goldberg, Informed Decision Making in Maternity Care, 18 J. PERINATAL EDUC. 32, 34 (2009).

\(^{112}\) RIMA D. APPLE, PERFECT MOTHERHOOD: SCIENCE AND CHILDBEARING IN AMERICA 161 (2006). Apple cautions, however, that “we must be careful not to romanticize this modern partnership of mother and physician.” Id. at 168. It has “created a new clinical world for both patient and doctor, a world in which there are no simple rules or procedures. Cooperation between mothers and experts should be our goal. But it will not be easy to attain.” Id. This success story explanation might be further supported by the increased role of women in obstetric care and the role of choice in health services as women “shop around” for the “right doctor.” MILLER, supra note 83, at 74, 77-78 (describing how women’s narrative accounts position the selection of hospitals, doctors, and pain relief protocols as means of gaining or retaining control). But see ANN BOULIS & JERRY A. JACOBS, THE CHANGING FACE OF MEDICINE: WOMEN DOCTORS AND THE EVOLUTION OF HEALTH CARE IN AMERICA 152 (2008) (concluding that “although differences in practice styles between male and female physicians exist . . . [s]ocial and structural factors will ultimately restrict such gender-linked differences”). Women are actively seeking out hospitals and doctors in ways that
omy has become the dominant and controlling principle in both informed consent law and medical ethics.”113 Modern women’s relationships with their medical experts are normatively framed more as a partnership, “albeit an unequal partnership,” whereby women work with their medical caregivers and, in turn, practitioners seek to understand their patients’ needs and encourage patients to bring questions and be informed.114 The American Medical Association acknowledges that “[t]he patient’s right of self-decision can be effectively exercised only if the patient possesses enough information to enable an informed choice” and that “the patient should make his or her own determination about treatment.”115

Women still value and retain a strong appreciation of science and medical expertise, but they also inform themselves and supplement medical guidance.116 Women have more access than ever to information about childbirth. They rely on books, the Internet, and relationships to supplement the expertise of their doctor, providing a diverse array of perspectives.117

Yet importantly, not all women are attaining such idealized partnerships. Rather, many women are not “afforded the ability of shopping for a doctor that will honor their beliefs.”118 Class, immigration status, insurance coverage, and geography reveal that this explanation cannot entirely account for the absence of conflict in medical decision-making. “Poor women are constrained by the conditions under which they have babies and the kind of care open to them . . . and this affects their ability to acquire knowledge about birth and their ability

assert their control over childbirth. Today, nearly half of the students enrolling in medical school are women. See id. at 2. (comparing this statistic to the eleven percent of women medical students in 1970). The authors caution that “complete integration remains elusive.” Id. at 190. Women are particularly strongly represented in pediatrics and obstetric/gynecology practice areas. See id. at 66 (noting that women represent 53% of pediatricians and 34.3% of OB/GYNs).

113. Benjamin Moulton & Jaime S. King, Aligning Ethics with Medical Decision-Making: The Quest for Informed Patient Choice, 38 J. L. MED. & ETHICS 85, 87 (2010). “[A]bsent the pregnant woman’s consent, her doctor has no more right to adopt the fetus as his ‘second’ patient than he does to make any of her other living children, or even her husband, his patient.” Oberman, supra note 4, at 473.

114. APPLE, supra note 112, at 125, 139 (2006) (explaining how these shifting hierarchies were brought about by the Women’s Health Movement and reformers such as Grantley Dick-Reid’s work on Childbirth Without Fear and Fernand Lamaze’s focus on childbirth preparedness, as well as the transformative publication of Our Bodies, Ourselves).

115. Moulton & King, supra 113, at 87.

116. APPLE, supra note 112, at 153 (noting that they do not rely much on innate expertise).

117. Id. at 144. See generally MAY FRIEDMAN, MOMMYBLOGS AND THE CHANGING FACE OF MOTHERHOOD (2013) (documenting the diverse range of maternal experiences reflected in the “mamasphere” and the vast numbers of women who engaged in Internet “dialogue and interactivity” to share experiences).

118. Hoffman & Miller, supra note 1, at 289.
to act on such knowledge.” Poor women more often give birth in public hospitals where they face long waits and interact with innumerable “nurses,” “aides,” “clerks,” “nutritionists,” “social workers,” and doctors. These institutional and inter-personal obstacles impede the flow of information, lead to contradictory advice, and complicate autonomy. Poor women birthing in public hospitals “rarely reach a point at which they have sufficient knowledge to manipulate the system to obtain more influence over their childbirth.” As the 2010 Blueprint on Maternity Care Report concluded, in the current model of hospital-based maternity care “[t]he vision of engaged and empowered childbearing women and families at the ‘center’ of well-coordinated maternity care is largely unrealized at present.”

And even the exercise of autonomy requires careful study of the doctor-patient relationship because informed consent requires unbiased thorough counseling. This is particularly important given the anomalous distinction of childbirth where the doctor needs to present information regarding maternal risks and fetal risks. We generally endorse the principle of individual autonomy but it is harder to position in the doctor-patient relationship. Women do not hold total agency in childbirth, nor is that necessarily the goal. Thus, while the actualization of women’s autonomy might partly explain the absence of conflict, it is far from a complete or universal explanation.

Alternatively, do birthing women and doctors align with such normalcy because women acquiesce to medical expertise? This explanation aligns with a longstanding historical shift to the primacy of

119. Lazarus, supra note 36, at 26 (internal citation omitted).
120. Id. at 32.
121. Id. at 32-33, 39 (further noting that women are, in turn, frustrated by these information gaps, and they struggle to even communicate that dissatisfaction to caregivers).
122. Id. at 39.
123. Angood et al., supra note 35, at S35 (concluding that the modern system “does not engage consumers as partners and empower them to take an active role in coordinating their own care”).
125. Shultz, supra note 53, at 221.
126. See, e.g., LUPTON, supra note 19, at 154 (noting that, despite these movements, “recent commentators have pointed out that such a shift in discourse and practice has not necessarily liberated women to enjoy freedom and agency while in childbirth”).
127. MILLER, supra note 83, at 31 (revealing “ways in which . . . expert knowledge is not rejected or even particularly resisted, but rather engaged with and thereby reinforced”).
128. See Rebecca A. Spence, Abandoning Women to Their Rights: What Happens When Feminist Jurisprudence Ignores Birthing Rights, 19 CARDOZO J.L. & GENDER 75, 97 (arguing that all women do not enjoy “meaningful birthing rights” and that “feminist lawyers can and must play a part in developing a robust conception of reproductive justice that includes birthing women, centering and prioritizing the needs of those with the least access to reproductive freedom”).
doctors in reproductive decision-making. Women endure tremendous pressure to be “perfect mothers.” This involves a deep pressure to make decisions that do not negatively impact their children. It is child-centered and it relies on the role of experts. If women do not “do everything (which means availing herself of technological birth), the process is her individual responsibility, and ultimately she must be blamed if she does not have the perfect birth.”

In this context, even for women, it is harder to position autonomy in the doctor-patient relationship when we hire doctors because of their expertise. Women’s own accounts of childbirth confirm that they “seek out and prioritise what they see as expert knowledge.” In fact, women have “increased engagement with expert bodies of knowledge and practices” and report that such practices are “reassuring” and help them “allay fears around perceived risks.” This reliance on experts is part of a transitional process into motherhood whereby uncertainty is mitigated by risk avoidance: “[S]ecurity is maintained throughout this period of transition based on a relationship of trust in experts and the knowledge that appropriate and responsible preparation, which implicitly diminishes risk, is being undertaken.” After birth, many women subsequently “question their ‘expert’ preparation” and question experts, demoting the positioning of experts.

129. See EHRENREICH & ENGLISH, supra note 30, at 28-30; LEAVITT, supra note 26, at 191.
131. Horwitz, supra note 130, at 47.
133. Lazarus, supra note 36, at 25, 27 (internal citations omitted).
134. Shultz, supra note 53, at 221.
135. MILLER, supra note 83, at 48.
136. Id. at 74.
137. Id. at 61; see also id. at 72 (“Through engagement with the medical profession and the regular monitoring of their pregnancies the women could be seen to be preparing to become mothers, in appropriate ways, reducing risk and acting responsibly.”).
138. Id. at 61-62 (describing this subsequent questioning as part of becoming a mother and “regaining a sense of . . . self”).
Acquiescence to medical expertise in childbirth also often involves acquiescence to technology. Modern women rely less on social support and inter-generational guidance, and defer more to technological understandings of pregnancy and childbirth.139 This changes the calculus of deference to medical expertise by leaving women to ironically perceive “greater uncertainty and risk” when expert knowledge is ordinarily called upon to achieve more certainty and predictability.140

Acquiescence to medical expertise invokes historical skepticism, however. As Lupton concludes, “[w]omen’s deference to the ‘doctor knows best’ ideology may be related to the asymmetry of information between doctors and patients, socialized respect for professionals with specialized training and for men in general.”141 It puts doctors in a position to “preempt patient authority.”142 Women’s own accounts of childbirth question the autonomy of acquiescence. They express frustration that their doctors acted like “they, rather than their ‘patients,’ knew best.”143 Women who knew and trusted their attendants believed their attendants “acted in their best interests” and “remembered their care far more positively,” while those with poor interpersonal relationships reported less positive experiences.144

The acquiescence to medical expertise explanation is further complicated by the complexities of modern medical decisions. Physicians work within a complex web of forces that shape their own decision-making, including private insurers, federal programs, and hospitals administrations.145 While women’s subordination, autonomy, and acquiescence to medical expertise might explain some degree of women’s alignment with their doctors, these accounts are polarized and even demonizing at times. The next section explores a more complex and nuanced explanation.

B. The Shared Fetal-Focus Framework

Some women and their doctors align in decision-making by adopting a framework that always selects the outcome that minimizes any risks to the fetus, presumptively and universally subordinating risks

139. Id. at 49-51 (noting how this phenomenon has been described as “technobirth”).
140. Id. at 48 (noting that this has moral underpinnings grounded in “ ‘responsible’ motherhood”).
141. LUPTON, supra note 19, at 158.
142. Shultz, supra note 53, at 221.
144. See DAVIS, supra note 49, at 107.
145. Lazarus, supra note 36, at 40 (noting that medical services are provided by many genuinely caring medical providers, but within the confines of a profit-based system).
to the birthing woman. This explanation raises complex and pervasive issues regarding how we understand obstetric care, decision-making methodologies, birthing autonomy, and reproductive rights. As the next section explores, while this decision-making methodology might actualize the autonomy of the women who elect it, it normalizes a problematic standard of care that creates an illusion of autonomy for all women.

While shared medical decision-making is not unique to childbirth, it is particularly distinct when the decision-making lens is focused on a putative third party—the fetus. This is not exclusively a problem involving doctors or the state giving primacy to fetal interests. Rather, doctors and women patients both purport to function in a fetal-focused frame.

Doctors do acknowledge that they often act primarily to minimize fetal risks. Many care providers admit that they feel “legally (or morally) responsible for the fetus and as such may override the needs of the women in order to assist the fetus.” The focus on minimizing all fetal risks has been offered to explain the absence of adequate informed consent in childbirth as well. The unique presence of the fetus in birth might account for the seemingly lax manner in which the informed consent doctrine has been applied in such cases. In the circumstances of labor and birth, the mother's individual right to informed consent must be weighed against a heavier counterbalance—the newborn infant—which carries with it a heavy load of emotional and cultural force.


147. Sue Kruske et al., Maternity Care Providers’ Perceptions of Women’s Autonomy and the Law, 13:84 BMC PREGNANCY & CHILDBIRTH 4 (2013); see also Jamie Abrams, Distorted and Diminished Tort Claims for Women, 34 CARDOZO L. REV. 1955 (2013) (concluding that women are subordinated to fetuses in the dual patient model).


149. Ketler, supra note 5, at 1039-40.
Fear of costly litigation for fetal harms might also drive doctors to prioritize minimizing all fetal risks.\textsuperscript{150} Scholars have previously acknowledged how this fetal focus is problematic for women.\textsuperscript{151}

Notably, however, many women are also electing the decision-making framework that minimizes any risks to the fetus although the rationales and implications are quite different for women than for doctors. Many women’s own framings of childbirth have changed too.\textsuperscript{152} Childbirth is often understood as a sacrifice distinctly of women’s health in the name of perceptively or actually minimizing risks to their fetus. Many identify mother as “synonymous with sacrifice.”\textsuperscript{153} “[M]others are now being usurped in the public consciousness by their fetuses,”\textsuperscript{154} reflecting a “giant collective wish for perfect mothering.”\textsuperscript{155} This leaves “the stakes of motherhood . . . so high.”\textsuperscript{156} One anesthesiologist described her own delivery as follows:

I don’t really care about the birth experience like a lot of patients do—into soft lights, soft music garbage. For me it was getting a good baby. I’ve seen too many times where patients are so concerned about it being a lovely experience for them that this has overridden the desires for having a good baby and they put themselves and their birth experience in front of having a “good” baby come out and having the best care for that baby.\textsuperscript{157}

These cultural shifts continue into parenting too. Many women today deploy “intensive mothering” frameworks “focus[ing] on chil-
dren to the exclusion of a focus on one’s own concerns as an adult.”¹⁵⁸ Modern parenting is uniquely child-centered.¹⁵⁹ While specific choices and strategies for parenting exist, this child-focused frame of modern parenting is “widely shared and often unquestioned.”¹⁶⁰ Modern parenting is “virtually synonymous with worry” as parents seek to ensure that their children are “healthy—physically, mentally, and emotionally.”¹⁶¹ These modern anxieties distinctly position parents as more engaged in the “formative stage, and believe that children’s experiences during the first two or three years of life mold their personality, lay the foundation for future cognitive and psychological development, and leave a lasting imprint on their emotional life.”¹⁶²

This approach—the shared focus on minimizing all fetal harms—is critical to examine when understood through a tort law lens, to understand how this framework is problematic for the women who do not adopt it. These dominant practices become “ritual[ized]” and then “transmit and reinforce gendered values”¹⁶³ that are, in turn, enshrined in tort standards of care, as explored below.

IV. WHY WHAT “MOST WOMEN” DO IS CRITICAL TO AUTONOMY FOR ALL WOMEN

The question of what “most women” do is deeply antithetical to reproductive rights advocacy. Women’s reproductive rights advocacy has worked extensively to defend the childbirth choices and autono-

---

¹⁵⁸. MARGARET K. NELSON, PARENTING OUT OF CONTROL: ANXIOUS PARENTS IN UNCERTAIN TIMES 19 (2010) (noting how changes in technology such as baby monitors and GPS systems have changed parenting greatly).

¹⁵⁹. Diane M. Hoffman, Power Struggles: The Paradoxes of Emotion and Control Among Child-Centered Mothers in Privileged America, in PARENTING IN GLOBAL PERSPECTIVE, supra note 132, at 229 (describing how modern parents focus on the child’s developmental needs and keenly respect the child as an individual).

¹⁶⁰. Id. at 230.

¹⁶¹. APPLE, supra note 112, at 1. Beginning in the 1970s distinctly, “parental anxieties greatly increased both in scope and intensity” as parents first sought to protect children from harms more consciously with inventions such as car seats, bike helmets, and babyproofing products. Parenting, ENCYCLOPEDIA OF CHILDREN AND CHILDHOOD IN HISTORY AND SOCIETY, http://www.faqs.org/childhood/Me-Pa/Parenting.html (last visited Feb. 10, 2015) [hereinafter Parenting]. This stands in stark contrast to earlier framings of American parenting, which have shifted from “adults in training” models to scientific models to the quest for emotional and psychological fulfillment. Id. The term “parenting” itself is new— injected with deep “critiques of value, practice and ideals and critiques of power.” PARENTING IN GLOBAL PERSPECTIVE, supra note 132, at 8; see also id. at 2 (“[P]arenting” also demands a discussion of reflexivity and individual ‘identity work’: to parent is to be discursively positioned by and actively contributing to the networks of idea, value, practice and social relations that have come to define a particular form of the politics of parent-child relations within the domain of the contemporary family.”).

¹⁶². Parenting, supra note 161.

¹⁶³. Miller, supra note 83, at 59.
my of women, particularly where such rights are positioned in conflict with their doctors, with the state, and with the fetus.164

But tort law reveals what “count[s]” as an injury in our society and which injuries matter more.165 Tort law is not an “objective system of adjudication”; rather, value judgments are embedded within this system to distribute suffering.166 Tort law does not just recognize and compensate injuries; it “does the political and social work of determining what will count as an injury and, ultimately, how it will be distributed.”167

It is thus critical that tort law standards are grounded in community-based determinations of reasonable behavior that are entirely shaped by what “most women” do. Thus, when a “community shares a value widely” that dominant value can become the standard of care.168 For example, many states approach informed consent from the perspective of what is significant to the “reasonable patient,” effectively “most patients.”169 For example, one informed consent birthing form states, “[f]etal monitoring by electronic machine is welcomed by the majority of mothers—any fears or questions?”170

Communities in tort law are distinctly invoked to make negligence law palatable. Community-based standard setting helps to “soften the hard surface” that the imposition of objective standards on indi-

164. See, e.g., Amy Kay Boatright, State Control Over the Bodies of Pregnant Women, 11 J. CONTEMP. LEGAL ISSUES 903 (2001) (examining the state’s authority to control a woman’s body during her pregnancy); Beth A. Burksrud-Reid, The Invisible Woman: Availability and Culpability in Reproductive Health Jurisprudence, 81 U. COLO. L. REV. 97 (2010) (examining how courts use the theoretical availability of alternative reproductive health services to prove that women’s health will not suffer and that courts also blame women for the lack of available services in ways that undervalue women’s health); V. Chandis & T. Williams, The Patient, the Doctor, the Fetus, and the Court-Compelled Cesarean: Why Courts Should Address the Question Through a Bioethical Lens, 25 MED. & L. 729 (2006) (presenting a bioethical lens to address conflict); Law, supra note 20, at 361-62; Nancy K. Rhoden, The Judge in the Delivery Room: The Emergence of Court-Ordered Cesareans, 74 CALIF. L. REV. 1951 (1986); Benjamin Grant Chojnacki, Note, Pushing Back: Protecting Maternal Autonomy from the Living Room to the Delivery Room, 23 J.L. & HEALTH 45 (2010) (proposing changes to promote maternal autonomy).

165. SARAH S. LOCHLANN JAIN, INJURY: THE POLITICS OF PRODUCT DESIGN AND SAFETY LAW IN THE UNITED STATES 13 (2006) (“[E]normous amounts of discursive energy frame and consolidate what will count as rational behaviors and whose interests these will privilege.”).

166. Id. at 34 (explaining how law functions in “highly specific contexts” and reflects the socially constructed view of “acceptable relations between persons and things”). Tort law “redistribute[s] human wounding . . . with vast implications of whose bodies the costs of progress fall into.” Id. at 5.

167. Id. at 2 (chronicling differences in legal responses to different types of product injuries).


169. King & Moulton, supra note 124, at 430.

170. KITZINGER, supra note 51, at 91.
viduals can create.\textsuperscript{171} Communities can stifle resentment of the more “distant, impersonal commands of negligence doctrine.”\textsuperscript{172} It allows the law to “outsource” the liability question to a group and “away from an abstract universal ideal.”\textsuperscript{173}

The role of community-based consensus among doctors is distinctly acute in setting standards of obstetric care. Tort law standards “require that physicians provide reasonable care under the circumstances, as judged against the level of knowledge and skill exercised by their professional peers.”\textsuperscript{174} Tort law gives a heightened deference to the customs of the medical community.\textsuperscript{175} Obstetric medical practitioners themselves set the standards of care that govern obstetrics, thus valuing collective professional medical organizations and consensus heavily. This standard setting and the uniquely community-based approach in which it occurs reveal how critical the tort law lens is to understanding the treatment of birthing women.

What “most women” do is also important because it shapes our very understanding of injuries. We process injuries against a larger social and political backdrop in a process that is “largely nonconscious or preconscious.”\textsuperscript{176} “[I]njuries are not objective facts; rather, they are events that humans perceive and interpret within ideational frameworks that reflect a deep interaction between self and culture.”\textsuperscript{177} They are processed in the context of the physical environment in which they occur.\textsuperscript{178} The processing of an injury is subject to recursive interactional influences of friends, family, and others “that takes place over time and draws third parties into the victim’s processes of cognition and response to injuries.”\textsuperscript{179}

\begin{itemize}
\item \textsuperscript{171} Anita Bernstein, \textit{The Communities That Make Standards of Care Possible}, 77 Chil.-Kent L. Rev. 735, 736 (2002).
\item \textsuperscript{172} \textit{Id.} at 739.
\item \textsuperscript{173} \textit{Id.} at 741.
\item \textsuperscript{174} \textit{Id.} at 764.
\item \textsuperscript{175} \textit{Id.} (explaining that, whereas for “many other occupations [courts] think of custom as merely relevant or admissible, the law of medical malpractice equates custom or substantially accepted practice with the standard of care”). Historically, this deference was quite problematic because the medical community perpetuated a norm of “cohesion” that prevented physicians from testifying against each other in court. \textit{Id.} at 764-65.
\item \textsuperscript{176} Engel, \textit{supra} note 81, at 303-04; \textit{see also id.} at 296, 321; \textit{id.} at 328 (explaining how “[t]he embodied mind would integrate [an injury] instantly and nonconsciously into its life story, and the injury victim would very likely describe the injury to others, including friends, family, and co-workers, as well as professional service providers,” and with each retelling arises an “opportunity for revision”).
\item \textsuperscript{177} \textit{Id.} at 319.
\item \textsuperscript{178} \textit{Id.} at 314-18.
\item \textsuperscript{179} \textit{Id.} at 306; \textit{id.} at 328 (describing how with each retelling of an injury “each listener might offer comments or reactions that alter the original perception and, recursively, help to create a revised narrative the next time around”).
\end{itemize}
The risk of bias is particularly problematic in the tort system. The tort system involves jurors comparing the conduct of stakeholders to their own “prototypes of how reasonable people behave.”\footnote{180}{Neil R. Feigenson, *The Rhetoric of Torts: How Advocates Help Jurors Think About Causation, Reasonableness, and Responsibility*, 47 Hastings L.J. 61, 73 (1995).} Jurors can employ many biases in this process. Accordingly cognitive bias complicates autonomy in cases where women seek to exercise their autonomy in ways other than that which most reduces risks to the fetus.\footnote{181}{Id. at 85-87.}

So, while antithetical to conventional framings of women’s birthing rights, the lens of what “most women” do is critical to understanding the standard of care that is applied to all women in childbirth.

V. HOW DECISION-MAKING MODELS TO MINIMIZE FETAL HARMs CREATE AN ILLUSION OF AUTONOMY

While women’s alignment with their doctors to choose the outcome that minimizes all fetal risks can actualize autonomy in individual cases, that approach—when adopted by communities of women—risks becoming the standard of care that is applied to all women. The framework is problematic because it creates an illusion of autonomy given the ease with which doctors can “preempt patient authority.”\footnote{182}{Shultz, supra note 53, at 221.} Women’s autonomy can easily be overridden by the doctor unilaterally selecting the outcome that best minimizes fetal risks and thereby foregoing actual informed consent and consideration of maternal risks. This positions doctors with a “trump card” to play to ensure that maternal-doctor conflicts arise only rarely. For example, “if you say to a woman that there’s a 1% chance this may save the baby’s life, she’ll take it.”\footnote{183}{Kitzinger, supra note 51, at 77 (internal citations omitted).} This is particularly so in the context of cesarean sections: “[m]ost women, if told by an obstetrician that a cesarean is best for the baby, go along with professional advice.”\footnote{184}{Id. at 80 (explaining how this accordingly becomes a “quick fix” to problems).} Thus, the model looks at first glance like it actualizes women’s autonomy, but it is an illusion because her decision is pre-ordained by the doctors’ communication of fetal risks and can be easily over-stepped, ignoring maternal risks. It suggests that medical providers’ “conscious belief in women’s autonomy may not translate to actual practice.”\footnote{185}{See Kruske et al., supra note 147, at 4.} Providing increased information to obstetric patients is not likely to change the outcomes.\footnote{186}{Law, supra note 20, at 365.} Thus, “all but the most idiosyncratic patients will agree with the doctor’s recommendation.”\footnote{187}{Id.}
The illusion of autonomy can be seen in one medical text that purports to instruct obstetricians in responding to conflict with birthing women over medical decision-making. *Clinical Obstetrics: The Fetus & Mother* includes a chapter on the medico-social considerations of pregnancy, which discusses the “Ethical and Legal Dimensions of Medicine of the Pregnant Woman and Fetus.” The text explains that any balancing between fetal benefit and maternal risk “must recognize that a pregnant woman is obligated only to take reasonable risks of medical interventions that are reliably expected to benefit the viable fetus or later child.” It states that, unbelievably, “[s]uch conflict is best managed preventatively through the informed consent process as an ongoing dialogue throughout a woman’s pregnancy augmented as necessary by negotiation and respectful persuasion.”

This illusion of autonomy can also be seen in the results of an Australian study considering health care professionals’ perceptions of women’s accountability and the providers’ own legal accountability. Notably, both midwives and doctors had previously agreed that women hold the right to autonomy in birthing. Participants were asked about the extent to which they agreed with this statement: “In collaborative practice, working with primary carers, the final decision should always rest with the woman.” Midwives agreed with this statement significantly more than doctors. When asked to rate disagreement with the statements, “For the safety of the baby, the maternity care team sometimes need[s] to override the needs of the woman’ and ‘Encouraging women to have more control over their childbearing compromises safety,’ ” doctors agreed that they sometimes had to override the woman’s interests, but midwives were more neutral with respect to that statement. Midwives disagreed significantly more than doctors with the idea that autonomy created safety concerns. Thus, doctors perceived a tension between autonomy and practice. As one researcher summarized, both midwives and obstetricians “only support women to make the final decision about an aspect of their care when this decision is what the care provider prefers.”

---

189. *Id.* at 1042.
190. *Id.* (emphasis added).
191. See Kruske et al., *supra* note 147.
192. *Id.* at 3.
193. *Id.*
194. *Id.* (emphasis omitted).
195. *Id.*
196. *Id.* at 4.
decision-making power” but then described the patients’ desires as “frivolous and as potentially conflicting with what [the doctor] decides ‘needs to happen.’”197 “Obstetricians continually described their primary role as sentries on the lookout for peril; they encountered conflicts as they heroically attempted to persuade misinformed or poorly informed patients to understand that doctors were the best judges of what is safe.”198 This is the illusion of autonomy.

Medical providers reported “a poor understanding of their own legal accountability, and the rights of the woman and her fetus.”199 This lack of clarity is problematic both for both women and their doctors. Health care professionals need clear guidance on how to handle requests to refuse medical treatment specifically in a system that values evidence-based decision-making and autonomy.200

Technology can distinctly perpetuate the illusion of autonomy in its perceived objectivity, but its distinct focus is on fetal outcomes. Fetal monitoring technology emerged in the 1960s to project images of the fetus, “seemingly independent” from the woman and “taking on a human form.”201 Notably, its scientific reliability and its necessity have been heavily contested.202

Fetal monitoring technology has “displaced the woman’s testimony from the central position it once held in the understanding of the fetus and the development of her pregnancy.”203 While a woman’s sensory awareness of her fetus, which long predated technological capabilities, used to be valued, many would say that today it is “completely ignored.”204 Fetal imagery perpetuates the illusion of autonomy because “representation of the fetus in isolation, abstracted from the body of the woman within which it is located, facilitates a perception of the fetus as a being that deserves no fewer rights than the wom-

197. Simonds et al., supra note 91, at 219.
198. Id. at 218.
199. Kruske et al., supra note 147, at 1.
200. Id.
202. Manual of Obstetrics, supra note 69, at 408 (concluding that “[s]ome form of evaluation of fetal well-being during labor is recommended,” but noting that “no randomized controlled study has ever shown continuous electronic fetal monitoring to be associated with better fetal outcomes than other forms of monitoring”). The authors explain that “[t]he main risk of electronic fetal heart rate monitoring is inaccurate pattern interpretation, thereby allowing a nonreassuring fetal status to go unrecognized or, conversely, and more commonly, precipitating unnecessary intervention in a healthy fetus.” Id.
204. Id. at 258, 264 (“It is not surprising, therefore, that the ultrasound image has been accepted as a virtually unchallengeable source of authoritative knowledge about the fetus, by professionals and laypeople alike.”).
It produces an “image of the fetus . . . independent of the uterus and sustains it, and by constituting the fetus as a patient in its own right, ultrasound has divided . . . the fetus from the pregnant woman.” So, the use of fetal technology pushes the fetus’s status as a patient forward, removes the woman’s voice as a patient (to varying degrees), and yields a primacy in the “objective” nature perceived to derive from fetal monitoring. “The animation of fetal life through such imagery did more than just personify the fetus. As the fetus emerged as a person, the pregnant woman began literally to disappear from view.” Indeed, fetal monitoring often triggers (many times inaccurately) fetal distress, which is the basis for cesarean delivery.

Yet, again, it is not exclusively a problem with medical professionals. While women have deferred to the medical expertise of doctors in childbirth for over a century, modern women distinctly defer to technological understandings of pregnancy and childbirth. For example, many women describe fetal imagery as “reassuring.” The reliance on medical experts reveals the “seduction of formal, medicalized preparation” for childbirth, seduction rooted in women’s “notions of risk, safety,” and desire to be “seen to act responsibly.”

It is hard—if not impossible—for women to “counterpoise the natural against the artificial, our intuitive, direct knowledge of our own bodies against the alien information derived from a machine.” This is an example of “demythologizing,” whereby women are not just trying to disprove an outcome, but to break an entrenched stereotype,

---

205. Morris, supra note 201, at 63.
206. NICOLSON & FLEMING, supra note 203, at 258, 264, 267 (explaining how fetal imagery “has been used beyond objective diagnostic tools” and explaining that the fetal image holds “affective, ethical, and religious rather than narrowly diagnostic” roles).
207. Id. at 262.
208. DANIELS, supra note 72, at 21.
209. STRONG, supra note 7, at 178.
210. “The concept of scientific motherhood . . . permeated . . . the country,” however, by the turn of the twentieth century, as mothers began to accept a “crucial role of contemporary science and medicine” in child rearing and care and “expect that medical and scientific experts and expertise should intervene in their daily lives.” APPLE, supra note 112, at 10, 33. Women actively sought out expert medical and scientific guidance in “all areas of child care, from mundane tasks to critical illness,” and this quest still thrives today. Id. at 33; see also id. at 157 (stating that popular culture reveals “just how far scientific motherhood [has] been normalized”).
211. MILLER, supra note 83, at 49 (noting how this phenomenon has been described as “technobirth”).
212. LUPTON, supra note 19, at 157 (explaining how women’s views of fetal imagery are shaped by their “age, past reproductive experiences, ethnicity, sexual preference, health status and desire for a child”).
213. MILLER, supra note 83, at 75.
214. NICOLSON & FLEMING, supra note 203, at 263 (explaining how we are perceived to “aid and improve our senses with technology”).
which can be futile. Doctors can easily cut the birthing woman out of the decision-making metric and act directly in the interests of minimizing fetal harms. As the former chairman of the Department of Obstetrics and Gynecology at Columbia-Presbyterian Medical Center in New York and co-chair of the American College of Obstetrics and Gynecologists’ committee reviewing obstetric practices once concluded, “we were beginning to forget that instruments such as electronic fetal monitors were tools to be used by the doctor, not decision-making machines to replace medical judgment.”

This illusion of autonomy can—and indeed has—distorted the standard of care itself, thus implicating all women, as explored more below.

VI. WHEN THE DECISION-MAKING FRAMEWORK OF MOST WOMEN IS APPLIED TO ALL WOMEN IT PROBLEMATICALLY ALTERS MEDICAL STANDARDS OF CARE

The minimization of all fetal harms decision-making model is problematic because its replication by “most women” in childbirth risks distorting the standard of care governing childbirth for all women. It suggests that the standard of care in childbirth requires complete compliance with medical advice and the minimization of all fetal risks. The actual standards of care, however, would require patient autonomy and only the minimization of unreasonable risks. This reveals the ghost of Roe v. Wade’s medical model.

A. Minimization of All Fetal Harms Instead of Unreasonable Risks

Standards of care are ordinarily framed around unacceptable or unreasonable risks. Yet the risk-avoidance behaviors discussed above suggest an elevated standard of care that requires the elimination of any risk to the fetus. This is socially constructed risk avoidance.

Indeed some doctors do suggest that the “minimization of fetal risks” is the lens that should govern women’s decision-making. One explicit example of this thinking can be seen in the specialty text, Ethics in Reproductive and Perinatal Medicine. The text acknowledges the bodily integrity of the birthing woman as an “important ethical value” crucial to the right of “self-determination.” It advises only “the most compelling of reasons” would allow for these rights to be

215. DANIELS, supra note 72, at 100.
216. Lazarus, supra note 36, at 28.
218. STRONG, supra note 7, at 180 (explaining that these rights are supported by a wealth of literature).
violated.\textsuperscript{219} It concludes that, while prevention of harm to the fetus is a “serious concern,” it cannot override the “normative personhood status” of the woman carrying the child.\textsuperscript{220} Yet, notably, the text then acknowledges repeatedly that the argument can be made that the woman has an “obligation to promote the interests of her fetus” and that this obligation increases in advanced gestational states.\textsuperscript{221} It states that mothers owe “an obligation to protect the offspring from harm.”\textsuperscript{222} These statements suggest poignantly that some degree of conflict is avoided by a wholesome presumption—the imputation of an unwritten duty even—that women will always act to minimize fetal risks.

\textbf{B. Complete Compliance with Medical Advice Has Never Been a Standard of Care}

It further suggests that complete compliance with expert advice becomes the standard of care by which women are judged to be “acting responsibly and avoiding unnecessary risks.”\textsuperscript{223}

During pregnancy, childbirth and motherhood, avoiding risk, and so being seen to be responsible, continues to involve placing trust in experts. To resist such engagement, to avoid screening tests, clinic visits and expert advice would be regarded as irresponsible behaviour. Such actions would be seen to jeopardise the woman’s own health “and more importantly, that of the foetus she is carrying and expected to protect and nourish in a proper maternal manner.”\textsuperscript{224}

As one woman articulated, “[t]here comes a point where you feel not trusting your doctor is not trusting your own judgment because you put time into selecting him and, should you begin to doubt him, you lose confidence in your own ability to make sound judgments.”\textsuperscript{225}

The “reasonable patient” matters greatly to tort law from the perspective of juror perception, comparative negligence claims, and informed consent models. Thus, the idea that all women are normalized toward a particular decision-making framework marginalizes those who adopt a different framework. Doctors’ accounts of patient inter-

\textsuperscript{219} \textit{Id.}
\textsuperscript{220} \textit{Id.} at 181.
\textsuperscript{221} \textit{Id.} at 179 (“Moreover, if there is going to be a future child with normative personhood status, the woman has a parental obligation to avoid actions that would be harmful to the child.”).
\textsuperscript{222} \textit{Id.} at 62.
\textsuperscript{223} MILLER, supra note 83, at 74, 87.
\textsuperscript{224} \textit{Id.} at 48-49.
\textsuperscript{225} Lazarus, supra note 36, at 38.
actions indeed describe “noncompliant patients as irritating and irrational.”

One doctor explains the discomfort of patient autonomy:

[Patient autonomy] makes it much more difficult. So I try to explain, in layman’s terms always, the consequences of the decisions and empower the patient to make the choice. Almost always they’ll end up choosing my recommendation when they realize that the choice is theirs. It’s very rare for someone, when they understand that my training says that [if] we go down this road, we do have the risk of compromising the baby, most folks choose and trust my training. Occasionally, when they don’t, it’s very difficult, but we can’t assault someone, you know!

Researchers describe how the obstetrician quoted above “saw her expertise as ultimately trumping patients’ contradictory viewpoints. She portrays women’s acquiescence as informed and sensible decision making rather than as an act of submission. As she sees it, patients must trust her to be the judge of whether what happens poses a risk to the baby.”

Requiring complete compliance with medical guidance would be distinct to childbirth, as any other patient can decline medical treatment.

It is only in the context of pregnancy that doctors assert the right to compel their patients to heed medical advice. Doctors’ responses to their pregnant patients therefore emerge as a startling exception to the nearly universal consensus that patients, not doctors, should control determinations about whether and when to undergo medical treatment.

This is particularly problematic when understood in conjunction with the preceding point whereby the standard is minimizing all fetal risks. Women must cede to medical authority at the expense of their own autonomy, but also often at the expense of their own medical risks. One author describes this as a “gestalt picture” where “[a]s the fetus comes into view, the woman disappears.” Thus, women have lost their autonomy and are being medically compelled to a self-sacrificial view of motherhood. “Good mothers, it is implied, should always wish to do what Doctor considers best for the fetus and unquestioningly take his advice.”

---

226. Simonds et al., supra note 91, at 218.
227. Id. at 217-18.
228. Id. at 218.
229. Oberman, supra note 4, at 469.
230. Morris, supra note 201, at 50.
231. Id. at 64.
This complete compliance with medical authority further reveals the resurrection of the ghost of Roe v. Wade’s medical model. The medical model in Roe positioned pregnant women as shared decision-makers with doctors, but really entrusted doctors with primacy. Roe squarely positioned the decision to terminate a pregnancy as a medical decision, and moreover, one in which the doctor distinctly held primacy over the pregnant woman. The Court held that, in the first trimester, the abortion decision “must be left to the medical judgment of the pregnant woman’s attending physician.” In the third trimester, the Court limited the state’s regulatory power by mandating an exception to prohibitions on abortion “where it is necessary, in appropriate medical judgment, for the preservation of the life or health of the mother.” Roe thus described the decision as one that “vindicates the right of the physician to administer medical treatment according to his professional judgment” and characterized the decision as “a medical decision” the “basic responsibility for [which] must rest with the physician.”

When read in conjunction with Doe v. Bolton, the medical lens was notably a broad one—at least initially and doctrinally—that positioned the physician to make decisions based on women’s health, age, family status, and emotional well-being. In Bolton, the companion case to Roe, the Court further reinforced and explained the medical frame of the decision to terminate a pregnancy. In Bolton, the appellants challenged the criminal abortion statute in Georgia. The statute had an exception de-criminalizing abortions when “continuation of the pregnancy would endanger the life of the pregnant woman or would seriously and permanently injure her health,” but it hinged that exception on the “best clinical judgment” of a physician. The plaintiffs challenged the statute, arguing that the “necessary” language was not objective enough and fearing that doctors would

233. This medical approach derived explicitly from historical advocacy leading up to Roe in which doctors and the American Medical Association sought to liberalize the criminal abortion laws by bestowing physicians with increased discretion to make the decision for their female patients in limited circumstances. Yvonne Lindgren, The Rhetoric of Choice: Restoring Healthcare to the Abortion Right, 64 HASTINGS L.J. 385, 387 (2013).
234. 410 U.S. at 164 (emphasis added).
235. Id. at 165 (emphasis added).
236. Id. at 165-66. Doctors had already exerted great influence in the early 1900s over birth control regulation, many opposing its legalization and ultimately securing a “monopoly over its delivery.” CARON, supra note 102, at 4-6 (articulating the implications of physician control over abortion on working-class women who could not afford the health care fees).
238. Id. at 181.
239. Id. at 183.
“choose to err on the side of caution and will be arbitrary.”\footnote{240} The Court upheld the district court’s holding that “health” was not vague and was “a judgment that physicians are obviously called upon to make routinely whenever surgery is considered,” explaining that the physicians’ medical judgment may be exercised in the light of all factors—physical, emotional, psychological, familial, and the woman’s age—relevant to the well-being of the patient. All these factors may relate to health. This allows the attending physician the room he needs to make his best medical judgment. And it is room that operates for the benefit, not the disadvantage, of the pregnant woman.\footnote{241}

\textit{Roe’s} progeny later modified this health exception in critical ways by limiting abortion to only the most extreme of medical harms facing the pregnant woman—functionally to save the life of the pregnant woman only.\footnote{242} The breadth of the medical framing in \textit{Bolton} is particularly noteworthy in the context of woman-doctor alignments. Notably, the “medical model” of abortion empowered doctors in reproductive decision-making and further empowered doctors to police their own professional membership.\footnote{243}

\textit{Roe’s} adoption of the medical model was widely criticized, yet it persists.\footnote{244} It is widely accepted that abortion jurisprudence has since shifted to the woman’s right to choose, away from a doctor’s right to decide in consultation with her. Yet, while abortion was characterized in \textit{Roe} distinctly in the context of health care, the health care delivery model for pregnancy termination became quickly isolated among specialist doctors in specialized medical facilities. Current framings of reproductive health “sever[] the right to decide to terminate a pregnancy from access to healthcare necessary to exercise that decision.”\footnote{245} The systematic access issues that this marginalization
creates are well documented and understood. Pregnancy termination services have been pushed out of health care models, and the physicians who provide the care have been pushed out of professional and social regard. The doctors who provide termination services have been demonized, marginalized, and ostracized from the medical profession through professional regulations, litigation, harassment, and violence.

The problematic bifurcation of obstetric health care and pregnancy termination health care was poignantly articulated in Gonzales v. Carhart. The Court concluded that doctors protect fetal interests, “abortion doctors” protect women’s autonomy, women are naturally destined to be mothers, and women who seek to terminate a pregnancy need to be protected. Indeed, Justice Kennedy in Carhart specifically reflected the magnification of medicalized fetal interests within the medical profession. His opinion reads as “[d]eeply skeptical that the medical profession has used the health exception in good faith,” and he “seems to believe instead that physicians have used the health exception as a proxy for promoting women’s autonomy at the expense of fetal life.” Carhart reveals this bias most poignantly, not only “sever[ing] abortion from healthcare, but also appear[ing] hostile to abortion providers,” describing them pejoratively as “abortion doctors” and suggesting that women need to be “protected from providers” who might fail to inform or guide them properly in their decision. This is a distinctly “woman-protective argument.”

The Court suggested that it was not just that women need to be protected from poor decision-making; it is that women who seek to exercise their decision-making autonomy in any way other than with a fetal focus are in need of protective barriers. It is a false choice for women: either act to protect fetal life or the state needs to protect you from your decision-making.

246. Id. at 390-91 (“Reclaiming abortion as a right of both healthcare and choice offers the potential for reclaiming the right within the larger framework of reproductive justice by granting all pregnant women, women who carry to term as well as women who choose to terminate their pregnancies, the right to exercise bodily autonomy and access healthcare in every aspect of their reproductive lives.”).


248. See id. at 159-60


250. Lindgren, supra note 233, at 405, 409-10 (explaining how Carhart and Casey have put “abortion exclusively as a right of choice, uncoupled from healthcare” (citing Carhart, 550 U.S. 124, and Planned Parenthood of Se. Penn. v. Casey, 505 U.S. 883 (1991)).

251. Id. at 410.

252. Id. (arguing that the “woman-protective argument conflates healthcare and choice: [a] woman must be protected from the abortion decision because the choice is harmful to her physical and mental health”).
While the medical model that bestows doctors with decision-making primacy over women has been largely abandoned in the context of pregnancy termination, this analysis suggests that its legacy persists and pervades obstetric care. Margo Kaplan concluded that Carhart’s “rationale can be imported into cases involving the medical treatment of women who wish to continue their pregnancies to term.” Carhart’s depiction of the “state interest in fetal life and maternal health” is cast “so broadly that it essentially creates new, dubious state interests that, in the context of compelled treatment cases, expand state justifications for requiring medical treatment of pregnant women, even where such treatment would harm women’s health.” This expanded reach of the state has problematic traction when understood in light of the modern framings of motherhood and parenting on which it feeds, as discussed above.

VII. RE-WRITING AND RE-RIGHTING OBSTETRIC STANDARDS OF CARE

Reproductive rights advocates have fought for decades to achieve women’s autonomy in childbirth, yet successes have not been universal or fully incorporated into clinical practices.

A. Methodological Standards of Care

To remedy the illusion of autonomy and the risks it poses to the standard of care, tort liability is needed for the dignitary harms created by the illusion of autonomy—methodological breaches in the standard of care. Incentive systems need to be changed in tort law to ensure that doctors are incentivized to emphasize how a decision is made, not just the substantive outcomes.

Doctors admit that methodological process is not prioritized. For example, one obstetrician quoted in a survey revealed the absence of methodological standards of care: “I think there are some patients who get very focused on the process, and all they care about is the process. . . . But for me, it is the end result. Do we have a good mom and a good baby? That’s what you try to do.” Researchers studying medical behavior and interviewing doctors explained how clients care about the birthing process, but doctors view their goal as “limiting

254. Id. at 145, 158 (“[Carhart] expands the ‘fetal life’ state interest far beyond what Roe and its progeny intended, essentially recognizing new state interests in promoting respect for human life and protecting women from medical decisions they might regret.”).
255. See, e.g., King & Moulton, supra note 124, at 429 (“Much has been written on how to bring the law to bear on medical practice in order to improve patient rights and protect physicians, but far less has been done to bring the practice of medicine to inform our legal standards.”).
256. SIMONDS ET AL., supra note 91, at 218.
bad things” and “achiev[ing] a good outcome.”

Doctors deprioritized a “process-related orientation” as less important than a “medicalized orientation” because the former can “jeopardize results, and concentrating on results justifies not attending to process.” In this medicalized framework, the “most important accomplishment is getting the baby out of the woman.”

Accordingly, methodological standards are needed to standardize the delivery of care. Because tort law has such strong “radiating effects,” it is critical to clearly articulate decision-making norms that can, in turn, directly shape and inform public discourse. Specifically, obstetric standards of care do not explicitly address how to resolve conflicts deriving from the simultaneous treatment of birthing women and their fetuses. Existing models are either nonexistent or not functional.

Informed consent only creates tort liability based on the actualization of negative outcomes, not the methodology itself. Informed consent examines whether a doctor negligently failed to disclose the nature, alternatives, risks, and consequences of a suggested treatment. The physician proposing treatment is required to inform the patient about all “material risks,” which include serious risks, even those with minute chances of actually occurring. The traditional standard focuses on what a reasonable medical professional would tell patients about the nature, alternatives, risks, and consequences of a given treatment. Some jurisdictions have adopted a standard in which the test is what “the reasonable patient would want to know.”

257. Id. at 219 (internal quotation marks omitted).

258. Id. (“Doctors did not represent birth as only pathological or risky, by any means, but they depicted it as always potentially pathological or risky. If you never know when disaster can strike, you must always be a sentinel.”).

259. Id. at 219, 222 (“[Within this discourse of OBs,] instrumental and operative deliveries are not conceptualized as risks to women, because risk is conceptualized as emanating from women’s bodies gone wrong or awry, rather than from acts done to women’s bodies by medical professionals.”).

260. See, e.g., Lisa Pratt, Access to Vaginal Birth After Cesarean: Restrictive Policies and the Chilling of Women’s Medical Rights During Childbirth, 20 WILLIAM & MARY J. WOMEN & L. 105, 121-22 (2013) (concluding that the notion that “all that matters is a healthy mom and healthy baby” does a disservice to women and babies because “women must be full participants in their pregnancies” so as not to “compromise the manner in which we attain the stated goal”). “Everyone agrees that healthy moms and healthy babies are important, but that goal must be carried out in a manner that acknowledges that the process is as important as the end result.” Id. at 122.


263. Id. at 196.

264. Id. at 195.

265. Id. at 196-97.
Modern reproductive rights arguments are being fought over the “winner take all” autonomy or primacy fight. Reproductive rights advocates have championed absolute women’s autonomy in reproductive decision-making, while opponents have argued for absolute personhood and fetal primacy. Analyzing the deeper social and cultural anchoring—as understood through the tort system—suggests reproductive rights advocates might be pursuing an unduly risky strategy by engaging the battle on these terms. It suggests that a methodological standard of care might be the most effective way to effectuate autonomy consistent with tort standards of care.

B. Using Decision-Making Aids

Tort standards of care should impose methodological standards of care for how decisions are made to ensure that the patient retains decision-making autonomy and that thoughtful, effective consultation is encouraged beyond traditional informed consent models. Decision-making aids should be required within the governing tort standard of care. Interactive decision-making aids facilitate effective decision-making and the processing of information.

Decision aids are tools that “collect and analyze the latest clinical evidence regarding the risks and benefits of different treatment options and then present the information in a manner patients can understand.” The process is to be collaborative with various stakeholders across various disciplines, including clinical researchers, practicing physicians, health services researchers, biostatisticians, and others, and they are regularly reviewed “to ensure both the accuracy and integrity of the information conveyed.” The decision aids provide information on the pros and cons of each option in an unbiased manner. In addition, the aids often offer video interviews and testimonials from patients and physicians regarding positive and negative experiences with each outcome and explanations for limita-

266. “[A]ntichoice activists have used fetal rights and mortality as their primary justification for restricting abortion . . . .” CARON, supra note 102, at 251.

267. See generally King & Moulton, supra note 124, at 480.


269. King & Moulton, supra note 124, at 464. See generally INFORMED MED. DECISIONS FOUND., FOUNDATION-FUNDED RESEARCH HIGHLIGHTS: ADVANCING OUR KNOWLEDGE OF HOW MEDICAL DECISIONS ARE MADE (Jan. 2013), available at http://informedmedicaldecisions.org/wp-content/uploads/2013/01/Research_Highlights.pdf (compiling all of the research the grants of the Informed Medical Decisions Foundation have supported over the last decade as they have worked to measure the problem, assess the quality of medical decision-making, and measure decision quality).

270. Angood et al., supra note 35, at S38.

271. King & Moulton, supra note 124, at 464.
tions in evidence for one treatment over another. Patients are given ample time to then digest and process the information and to make their communications with their physicians more fruitful.

Childbirth is distinctly well-positioned for decision-making aids because there is time to plan and prepare. It is also cost-effective to create decision aids for childbirth because the decisions are repeated so many times and so consistently. In fact, some decision aids already exist for certain medical procedures.

Decision aids are currently being integrated into various legislative and clinical settings. The Federal Affordable Care Act, for example, makes grants available to health care providers “for the development and implementation of shared decision-making techniques and to assess the use of such techniques.” Various medical centers in the United States have started experimenting with decision-making aids. A Washington statute also requires competent pa-

---

272. Id.

273. Id.

274. See Sakala & Corry, supra note 23, at 66 (“[P]regnant women have many months to prepare and would benefit from high-quality information and decision support relating to labor and birth well before labor.”).


276. King & Moulton, supra note 124, at 465.

277. Affordable Care Act, 42 U.S.C. § 299b-36(e)(3)(A) (2010). See generally id. § 299b-36 (stating that the purpose of this section is to “facilitate collaborative processes between patients, caregivers or authorized representatives, and clinicians that engages the patient, caregiver or authorized representative in decision making, provides patients, caregivers or authorized representatives with information about trade-offs among treatment options, and facilitates the incorporation of patient preferences and values into the medical plan”); Samuel F. Hansen, The Role of Decision Aids in the Affordable Care Act, STAN. J. PUB. HEALTH (2013), available at http://web.stanford.edu/group/sjph/cgi-bin/sjphsite/the-role-of-decision-aids-in-the-affordable-care-act/. Section 3506 of the Affordable Care Act, “Program to Facilitate Shared Decision-making,” provides standards for the developing, implementation, funding, and certification of decision aids within the national health care system. Patient Protection and Affordable Care Act, Pub. L. No. 111-148, § 3506, 124 Stat. 119, 527 (2010). The section facilitates decision aids by funding a third party entity which will develop standards based on medical consensus and certify patient decision-making aids for use by federal health programs. Id. at 527-29. Patient decision aid is defined as “an education tool that helps patients, caregivers or authorized representatives understand and communicate their beliefs and preferences related to their treatment options and to decide with their health care provider what treatments are best for them based on their treatment options, scientific evidence, circumstances, beliefs and preferences.” Id. at 527. The aids inform patients and their families of the benefits, risks, costs and effectiveness of tests and treatments. See Emily Oshima Lee & Ezekiel J. Emanuel, Shared Decision Making to Improve Care and Reduce Costs, 368 NEW ENG. J. MED. 6, 7 (2013).

278. Moulton & King, supra 113, at 91 (highlighting a Dartmouth-Hitchcock Medical Center for Shared Decision-Making); see also Informed Choice, DARTMOUTH INST., http://tdi.dartmouth.edu/research/engaging/informed-choice (last visited Feb. 10, 2015) (explaining that shared decision-making is needed when there are multiple choices, each
tients to sign written acknowledgement that she engaged in shared decision-making with a certified decision aid.279

The process of developing decision aids would also strengthen evidence-based standards of care. They would reveal the “critical gaps in the evidence needed for decision making on [specific obstetric decisions], then fund and conduct targeted research with time frames that can compare short-term and longer-term outcomes and costs.”280

Embedding a decision-aid framework within the standard of care by which doctors consult with their patients would help remedy the issues described here. This is a tool to facilitate the decision outcome, separate and distinct from informed consent, by which the doctor presents the risks and benefits. One study considered the efficacy in presenting a decision-aid framework to help women decide whether to have a subsequent caesarean birth after a prior cesarean delivery.281 The study concluded that such a tool was useful in helping women consider the risks and benefits of delivery options, and it improved the extent to which they felt informed, but it did not improve their relationships with their health care provider.282 High-quality decisions come from “a strategy of ‘vigilance’, where the decision-maker searches for information that is relevant to the decision, assimilates information in an unbiased manner and then appraises the alternatives before making a choice.”283 A workable decision-making framework would need to acknowledge individual habits that skew toward “unquestioned acceptance, responsibility shifting, rationalisation, bolstering of the least stressful alternative and inattention to additional information that would involve change.”284 Framing a decision-making model would account for the real-world realities of decision-making in the context of time and environmental constraints, such as family or social commitments, relationships, and risk averseness.285

with its own advantages and disadvantages, none of which is a clear “correct” choice, but rather the “correct” choice depends on individual factors and values).

279. Moulton & King, supra 113, at 92.
280. Angood et al., supra note 35, at S38.
281. Allison Shorten et al., Preparing Consumers for Shared Decisions: Analyzing the Effectiveness of a Decision-Aid for Women Making Choices About Birth After Caesarean, in PSYCHOLOGY OF DECISION MAKING IN HEALTH CARE 73 (Elizabeth P. Blakely ed., 2007) (studying 227 women). This is a complicated decision because both options are considered “safe” for most women, and both options involve some degree of risk for the mother and the baby. Id. at 78. The risk of uterine rupture is relatively small, less than one-half percent, but the consequences of the risk occurring are huge—hysterectomy or fetal death. Id.
282. Id. at 93-94.
283. Id. at 76.
284. Id.
285. Id. at 77, 79.
Decision aids better perpetuate a model of autonomy. They allow the patient to “make an autonomous choice to participate in a full or limited way or not at all in making the final decision after receiving the relevant information.” Thus, patients can elect to defer fully to medical judgment in ways that remain consistent with autonomy, but it is not per se the standard of care that they have to do so. In that sense, the aids are beneficial to doctors and patients alike. For doctors, decision aids would better address the accountability fears of medical practitioners who perceive conflicts between tort liability for bad fetal outcomes in obstetric care and actualizing women’s autonomy. Including a decision-making framework in the standard of care would better protect doctors when bad fetal outcomes occur and better protect women’s autonomy in the cases where women deviate from normalized mainstream decision-making methodologies.

Decision-making aids can be an effective tool to deter the cognitive bias created by the experiences of most women and most doctors in tort law. Cognitive bias researchers describe that several mechanisms can reduce or eliminate the problems of cognitive bias. The replacement of human intuition—the origin of cognitive bias—with formal procedures can strengthen decision-making frameworks. Many decision-makers lack appropriate “codes” to detect bias. However, this can be remedied by external calibrations. Researchers generally believe that it is possible to “debias” a problem; subjects when faced with “falsifying evidence . . . generally did reject their hypotheses . . . and respond accordingly,” “suggest[ing] that subjects may be passive rather than active.” In that sense, decision aids can also be a highly effective response to race, class, and ethnic differences in child care decision-making.

Yet the aids themselves need to be managed for risks of bias, which, of course, is extremely sensitive within the history of the state’s problematic regulation of informed consent in abortion. To

---

286. King & Moulton, supra note 124, at 436.
287. Id.
288. Jonathan St. B.T. Evans, Bias in Human Reasoning: Causes and Consequences 114-15 (1989) (explaining the general belief that critical thinking is a skill that can be taught, yet expressing skepticism regarding this accepted conclusion).
290. Evans, supra note 288, at 49-50 (noting, however, that the de-biasing mechanisms need to be more interactive than just instructions to decision-makers).
291. Angood et al., supra note 35, at S30.
292. Moulton & King, supra note 113, at 92 (describing a tension with how states have managed abortion informed consent and proposing resolution through effective certification procedures).
ensure that materials are unbiased, proper certification from neutral bodies is required.\footnote{293}{King & Moulton, supra note 124, at 466.}

Decision aids are also uniquely time and resource intensive to prepare as a tool.\footnote{294}{Id.} They are further burdensome on physicians to implement.\footnote{295}{Moulton & King, supra note 113, at 90 (explaining how it has been challenging to integrate decision aids in practice because of administrative challenges and a lack of financial and legal incentives).} Yet they can achieve critical improvements in “patient comprehension” and “decisional conflict,” and they can also achieve some “improved health outcomes.”\footnote{296}{Id.}

Successful implementation of decision aids requires compulsion through a tort-based standard of care. Absent a clear standard of care requiring such methodological precision, doctors lack the incentives and tools to implement.\footnote{297}{See id. at 92.} The standard of care would need to explicitly protect doctors who follow careful methodological decision-making from later litigation. For example, the Washington State bill treats the signed acknowledgment of shared decision-making with the use of a decision aid as prima facie evidence of informed consent.\footnote{298}{Id. Washington State positions shared decision-making as “an alternative means of meeting the informed consent requirement set forth by laws . . . .” WASH. REV. CODE § 7.70.060(2)(a) (2012).}

VIII. CONCLUSION

The complexities of obstetric care where dual duties are owed to both the fetus and the birthing women suggest that there should be more conflict in obstetric care. Many women and their doctors resolve this complexity by deploying a decision-making model to minimize all fetal risks. This standard is unworkable and intolerable because it perpetuates a mere illusion of women’s autonomy. The deference to medical determinations of fetal risks further resurrects the ghost of \textit{Roe v. Wade}’s medical model. It grossly deviates from baseline standards of care by exaggerating the severity of fetal risks and undervaluing even unreasonable risks to the birthing women. Cognitive bias further entrenches this implicit model and exacerbates it along class, race, and ethnic lines. Tort law should care about the methodology of decision-making, just as it cares about the substance. Furthermore, it should be used to create/install the methodology by which women can regain/establish their actual autonomy over medical decision-making.
PARTITIONING AND RIGHTS: 
THE SUPREME COURT’S ACCIDENTAL 
JURISPRUDENCE OF DEMOCRATIC PROCESS

JAMES A. GARDNER*

ABSTRACT

In democracies that allocate to a court responsibility for interpreting and enforcing the constitutional ground rules of democratic politics, the sheer importance of the task would seem to oblige such courts to guide their rulings by developing an account of the nature and prominent features of the constitutional commitment to democracy. The U.S. Supreme Court, however, has from the beginning refused to develop a general account—a theory—of how the U.S. Constitution establishes and structures democratic politics. The Court’s diffidence left a vacuum at the heart of its constitutional jurisprudence of democratic process, and like most vacuums, this one was almost immediately occupied. But the Court filled its jurisprudential hole not primarily by invoking principles of democracy—even unstated ones—but by doing instead what reluctant decision makers often do: by reaching for whatever is handy. In a path-dependent series of small but fateful steps, the Court’s reaction took two main forms. First, in the absence of a pertinent theory to guide it, the Court fell back on habit, specifically a habit, developed in its earliest cases, of solving problems of political power and representation by partitioning the electorate—that is, by ordering it subdivided. By resorting reflexively to this approach, the Court soon came to treat partitioning as the preferred solution to most problems of democratic representation. Second, the Court reached for the tools of decision that were most ready at hand, and those tools were individual rights, initially equal protection, then the freedoms of speech and association. But because these tools were ill-suited to the task, the Court ended up stretching First Amendment analysis in these cases beyond its plausible bounds and purposes. A well-ordered democratic state needs a thoughtful and deliberate jurisprudence of democracy and democratic practice. Instead, the Court has provided an accidental, haphazard jurisprudence of habit and availability.

I. INTRODUCTION

If the main function of a constitution is to set the ground rules by which a polity governs itself,1 then in constitutional democracies

---

* Bridget and Thomas Black SUNY Distinguished Professor, SUNY Buffalo Law School, The State University of New York. An earlier version of this paper was presented at a workshop on Electoral Law: The Virtuous or Vicious Circle of Theory and Practice, at McGill University Law School, Montreal, Quebec, November 19, 2013. I wish to thank the organizers, Hoi Kong, Han Ru Zhou, and Maxime St.Hilaire, for their kind hospitality. My thanks also to Guy Charles and Michael Halberstam for comments on a prior draft, and to Andrew DeMasters for valuable research assistance.

1. See, e.g., JOHN LOCKE, THE SECOND TREATISE OF GOVERNMENT, ch. X (Richard H. Cox, ed. 1982) (1690) (describing how members of a civil society create a form of government). For a more contemporary account, see RUSSELL HARDIN, LIBERALISM,
surely the most significant ground rules are those structuring democratic politics. These are the rules that establish the basic framework within which social disagreements are resolved, the processes by which binding agreements are negotiated, and the criteria by which such resolutions are to be deemed legitimate by those of whom submission to official power is demanded.2

In polities that, like the United States, allocate to a court responsibility for interpreting and enforcing the constitutional ground rules of democratic politics, the sheer importance of the task would seem to oblige such courts, when adjudicating disputes over basic democratic processes, to guide their rulings by developing an account of the nature and prominent features of the constitutional commitment to democracy. The very definition of a constitution is sometimes said to include not only the constitutional text, but also a “nation’s . . . dominant political theories.”3 It is widely agreed that courts cannot in practice decide constitutional cases involving regulation of the democratic process without resort to some underlying theory of democratic politics—“engagement with structural theories in election law is inescapable.”4 As Heather Gerken has explained in the context of redistricting, “[c]ourts cannot decide whether power has been ‘fairly’ or ‘properly’ allocated among voters without having a broader theory of how a healthy democracy should function . . . .”5 The high courts of other nations have not shrunk from developing such accounts, or at least from making a serious attempt.6


5. Gerken, supra note 4, at 521.

6. For example, according to Yasmin Dawood, the Supreme Court of Canada “has played an important role in defining Canadian democracy.” Yasmin Dawood, Democracy and the Right to Vote: Rethinking Democratic Rights under the Charter, 51 OSGOODE HALL L.J. 251, 253 (2013). The Australian High Court has inferred a freedom of political speech
Nevertheless, the U.S. Supreme Court has long refused to develop a general account of how the U.S. Constitution establishes and structures the democratic politics occurring within the very institutions that the Constitution itself creates: “Members of every generation of the Supreme Court’s Justices have claimed that they have no theory about the way democracy should work.”7 Until the mid-twentieth century, this refusal had few consequences because the Court did not understand its power to extend to policing the operation of democratic institutions.8 In the Court’s view, the Constitution did not subject democratic politics to judicially enforceable constitutional meta-rules, and for judges to attempt to find them in the Constitution exceeded the legitimate bounds of the judicial role by asking them not to apply law, but “to choose . . . among competing theories of political philosophy.”9

By 1962, however, the Court changed its view of its own powers, and began to intervene regularly—and with increasing impact—in the business of deciding the ground rules of democratic politics.10 At the same time, the Court continued to refuse to develop an account of the constitutionally grounded structure of democratic processes. As a result, the Court has over the last five decades decided cases substantially reshaping the political landscape—eliminating restrictions on voting,11 overturning long-established institutions of political representation,12 and invalidating regulatory limits on political speech and spending13—largely without a compass.


9. Baker v. Carr, 369 U.S. 186, 300 (1962) (Frankfurter, J., concurring). Among members of the contemporary Court, Justice Thomas has most explicitly expressed a similar sentiment: “[M]atters of political theory are beyond the ordinary sphere of federal judges. And that is precisely the point. The matters the Court has set out to resolve in vote dilution cases are questions of political philosophy, not questions of law.” Holder v. Hall, 512 U.S. 874, 901 (1994) (Thomas, J., concurring).

10. The pivotal case was Baker v. Carr, 369 U.S. 186 (1962). This history is elaborated infra Part II.


Scholars have advanced several possible explanations for the Court’s surprising reticence. It has been suggested, for example, that the Court avoids attempting to tease out of the Constitution some plausible baseline theory of how American democratic politics ought to work because the issues are so “hard to figure out”\(^\text{14}\) that the task may surpass judicial competence.\(^\text{15}\) Others argue that judicial development of such a theory is properly avoided because the Constitution’s indeterminacy on the subject of democratic practices raises unacceptable risks that courts might improperly ossify contingent political arrangements that are best left fluid, or that judges might rely excessively on their personal views of what democracy requires.\(^\text{16}\) Another family of explanations proposes the Court’s embrace of a minimalist approach to judging\(^\text{17}\) or the justices’ preference for highly specific doctrinal formulae couched at low levels of abstraction.\(^\text{18}\)

Although there may be a grain of truth to all these explanations, in the end they give the Court more credit than it deserves. A close and careful look at the precise sequence in which the Court’s jurisprudence of democratic process evolved tells a different story, one distinctly less appealing on account of the almost complete absence from the Court’s decision making of deliberate judicial choice and reflection. What this history shows is not the application of some consistent and coherent judicial philosophy or practice of judging; to the contrary, it shows that the Court’s jurisprudence of democracy arrived at its present unsatisfactory state accidentally, by way of a path-dependent sequence of small yet fateful steps.

Specifically, the Court’s lack of a theory of democratic politics in its earliest cases left a vacuum at the heart of the constitutional jurisprudence of democratic process. Like nature, however, jurisprudence abhors a vacuum; cases must be decided somehow, on some basis, if decisions are to be taken. I argue here that the Court filled its jurisprudential hole not primarily by invoking principles of democracy—even unstated ones\(^\text{19}\)—or by invoking and following consistently some set of beliefs about the modest role of courts in a democra-

\(^{14}\) See Gerken, supra note 4, at 508; Gerken, supra note 7, at 421.

\(^{15}\) See Kang, supra note 4, at 1099-100; Lowenstein, supra note 7, at 302.


\(^{17}\) See Kang, supra note 4, at 1105. On the risks of judicial grand theory, see Cass R. Sunstein, One Case at a Time: Judicial Minimalism on the Supreme Court (2001).

\(^{18}\) See Gerken, supra note 7, at 421-22.

\(^{19}\) See, e.g., Charles, supra note 4, at 1114; Gerken, supra note 7, at 414; Kang, supra note 4, at 1113 (all arguing that it is impossible for courts to decide cases dealing with democratic processes without orienting themselves against some underlying conception of democracy, which is thus necessarily present even if unarticulated).
cy. Instead, the Court responded to questions about democratic process raised in its cases by doing what reluctant decision makers often do: by reaching for whatever is handy. This reaction took two main forms. First, in the absence of a pertinent theory to guide it, the Court fell back on habit. Its principal relevant habit, developed in its earliest cases dealing with democratic practices, had been to solve problems of political power and representation by partitioning the electorate—that is, by ordering it subdivided, thereby converting minorities within a jurisdiction into local majorities in a smaller one. By resorting to this habit in subsequent cases, the Court soon came to treat partitioning as the preferred solution to most problems of democratic representation, even where it might be of dubious wisdom.

Second, the Court reached for the tools of decision that were most ready at hand, and those tools were individual rights. In particular, the Court unthinkingly imported an antidiscrimination approach, pioneered in cases involving racial discrimination and relying on principles of equal protection, into a large number of disputes dealing with democratic process, problems for which this approach often was not well-suited. Later, when the Equal Protection Clause began to prove inadequate to the increasingly complex task of regulating political processes, the Court began, reflexively, to import other rights into the democracy arena, principally the First Amendment protections of speech and association, which it proceeded to stretch badly by applying them in circumstances for which they were not designed. As a result, the Court has developed what might plausibly be called a jurisprudence of habit and availability when it should have developed a jurisprudence of democracy and democratic practice.

How the Court reached this point is the story I wish to relate. It begins not with modern, frustratingly complex conceptual problems of campaign speech and finance, but with very old problems of brute, physical territoriality arising from the way human beings distribute themselves on the land. Part I therefore reviews briefly the historical evolution of the American system of territorial representation, demonstrating how a set of institutions initially well-matched to pre-


21. See, e.g., Pamela S. Karlan & Daryl J. Levinson, Why Voting Is Different, 84 Calif. L. Rev. 1201, 1216 (1996) (“[T]he Court . . . has . . . created a doctrinal morass by selectively wrenching concepts out of the contexts in which they were developed and attempting to jury-rig them to work in a context where they do not make sense . . .”); Samuel Issacharoff and Pamela S. Karlan, The Hydraulics of Campaign Finance Reform, 77 Tex. L. Rev. 1705, 1705-06 (1999) (the Court’s practices have “led to the recasting of essentially political challenges born of electoral frustration as racial ones”).
vailing political theories and practices came increasingly under pressure as social and political beliefs evolved. Part II examines the U.S. Supreme Court’s entry onto this terrain in its early decisions reviewing the constitutionality of laws regulating democratic representation. Paying careful attention to the kinds of cases coming before the Court and their precise sequence, Part II shows how the Court’s earliest rulings, in cases involving racial discrimination in politics, established a template for judicial intervention that the Court, in path-dependent fashion, almost immediately applied in other, far less propitious settings.

Part III describes the quick emergence and solidification of the Court’s preference for solving problems of democratic representation by partitioning the electorate so as to transform complaining political minorities into content, locally dominant majorities. It argues that the Court’s habitual and unthinking reliance on partitioning, unmoored to any normative conception of who ought to be represented in legislative bodies or their appropriate degree of influence, led predictably to the emergence of one of the most intractable problems in contemporary American politics: the Court’s complete inability to adjudicate successfully questions of partisan gerrymandering.

Finally, Part IV examines the Court’s unguided deployment of generic individual rights to solve problems of democratic process and participation, following the Court’s jurisprudence step by step from an initial reliance on equal protection to its expansion into First Amendment freedoms of speech and association. Part IV shows how reliance on principles of equal protection, which had served the Court well in its early cases involving racial discrimination in politics, began to cause the Court mounting problems, leading it to turn to the freedom of speech, where it ran into even more severe and less tractable problems. Ultimately, constrained by a series of path-dependent decisions into penetrating ever more deeply into the use of individual rights—a commitment from which it apparently saw no possible retreat—the Court has continued to press First Amendment rights into service to a point well beyond the bounds of logic or necessity, a widely criticized pattern that continues to this day.

II. EVOLUTION OF THE AMERICAN SYSTEM OF TERRITORIAL REPRESENTATION

American institutions of political representation have their roots in an English system designed initially to represent land.22 Land-
holders in feudal England held their estates under an obligation to provide various forms of aid to the crown, including, upon request, financial assistance. Because financial impositions by tradition could not be assessed without the consent of those tenured in the lord’s land, representatives of the land were summoned to Parliament for the purpose of giving their consent to taxation.

As the rise of commerce expanded the potential sources of wealth beyond land, English monarchs sought to tap these new sources of royal revenue by expanding Parliament to include representatives from corporate towns and boroughs, where merchant wealth was mainly to be found. Nevertheless, representation in Parliament continued to be based on the unit from which consent was required, irrespective of its actual characteristics, including who or how many happened to live there, or even the amount of revenue due from the taxable unit. By the late fourteenth century, representatives in Parliament consisted of two knights from each county and two citizens or burgesses from each city or borough within the represented counties, regardless of population, wealth, or property value.

This model eventually crossed the Atlantic to the American colonies, where representation in colonial legislatures was allocated not to individuals, but to local communities. Thus, in Massachusetts, representatives represented towns; in Virginia, plantations, hundreds or counties; and in the Carolinas, parishes. As the Massachusetts Supreme Judicial Court said in 1811, “[t]he right of sending representatives [to the state legislature] is corporate, vested in the town . . . ” By the time this method of representation became entrenched in the colonies, however, its justification had evolved from one based on feudal obligations in land to a more characteristically republican justification that presupposed a commonality of interest arising from shared characteristics of the inhabitants of represented units:

The corporate method of representation presumed that physical proximity generated communal sentiment. Each geographic unit was thought to be an organic, cohesive community, whose resi-
dents knew one another, held common values, and shared compatible economic interests. The smaller the community, the more likely that its citizens would identify with one another . . . . Large distances, in contrast, bred a diversity of peoples and values.30

By the time of the Revolution, the founding generation fully accepted this account of representation. The idea that the political interests of communal groups of individuals correlated strongly with territory served, for example, as an axiom in Madison’s famous defense of the large republic in Federalist 10.31 “Factious combinations,” Madison argued, are “less to be dreaded” in a large republic than in a small one because of the greater variety of interests found among a larger populace, a characteristic that is entirely an artifact of geographical scale: “Extend the sphere and you take in a greater variety of parties and interests; you make it less probable that a majority of the whole will have a common motive to invade the rights of other citizens . . . .”32 The idea that territorially defined local communities may reliably serve as proxies for the shared, collective interests of the individuals who inhabit them has remained a fixture in American political thought ever since.33 So, for example, among delegates to the Wisconsin constitutional convention of 1851, “[t]he leading idea seems to have been that each county was regarded in the nature of ‘a small republic,’ or ‘in the light of a family,’ and ‘each organized county had a separate interest.’ ”34 In its more modern incarnation, the belief that place and interest coincide centers on the idea of the “community of interest,” a term widely used in federal reapportionment jurisprudence.35

Nevertheless, it is not immediately self-evident why the inhabitants of any particular locality should comprise a community of interest. Why should mere common habitation of a unit of local government reflect, or give rise to, a community of interest among the residents? The answer to this question was gradually worked out by American state courts, which typically offered two distinct, though not unrelated theories. First, the inhabitants of a county or similar

32. Id.
33. For a strong critique of this phenomenon, see Thomas Bender, Community and Social Change in America 4-8 (1978).
34. State ex rel. Attorney General v. Cunningham, 51 N.W. 724, 739 (Wis. 1892) (Piney, J., concurring), (quoting Journal of Debates 219-24 (1851)).
local government unit were said to share a common local economy and economic life; second, county residents were said to participate together in the public life of a shared unit of political and governmental administration.\textsuperscript{36}

The linkage in the state constitutional jurisprudence between counties as the basic constitutional unit of representation and the shared economic life of county residents has never been expressed more clearly than in a 1964 opinion of the New Jersey Supreme Court:

Anciently, and still today, the counties reflect different economic interests, although of course these economic interests are not perfectly contained or separated by any political line, municipal, county or State. So, certain counties have a dominant concern with manufacturing and commerce; others have a large stake in agriculture; still others lean heavily upon the resort industry; and finally a few counties have a special interest in the products of the sea.\textsuperscript{37}

The idea here, clearly, is that counties are not arbitrary territorial units, random shapes on a map, to be ignored or rearranged on a whim, but rather contain populations that have distinctive interests, and these interests are primarily economic; each county, that is to say, comprises a distinct local economy.

At the same time, American state courts also have frequently found that residency in a county creates what might be called an “administrative” community of interest among the inhabitants in virtue of their common experience of the county’s administration of governmental programs.\textsuperscript{38} Finally, state courts have sometimes found that common participation by a county’s inhabitants in its electoral politics, and in the reciprocal relationships established between those inhabitants and their elected officials, gives rise to a political community of interest entitled to recognition.\textsuperscript{39}

For much of American history the activities of state legislatures conformed largely to this model. During the colonial period, the matters that individual legislators brought to state legislature were “basically the business of their fellow townsmen,” and the legislative

\textsuperscript{36} This argument is worked out in greater detail in Gardner, Representation without Party, supra note 22, at 939.


\textsuperscript{38} On the role of counties in state government, see, for example, \textit{In re Legislative Districting of the State}, 805 A.2d 292, 319 (Md. 2002), and Stephenson v. Bartlett, 562 S.E.2d 377, 385-86 (N.C. 2002).

agenda was set essentially by petitions from towns and individuals.\textsuperscript{40} This changed surprisingly little over the course of the nineteenth century. During much of that period, state legislatures “spent most of their time responding to highly specialized demands like divorces or the settlement of local disputes and land titles.”\textsuperscript{41} In mid-nineteenth-century Maryland, for example, no more than ten percent of state legislation took up matters affecting the entire polity, whereas more than half of state laws affected only specific local communities and groups, and one-third provided some kind of benefit to specific individuals.\textsuperscript{42} Legislative politics was thus conceived primarily as an arena for satisfying demands made by communities and individuals, not as one for taking up universally applicable programmatic initiatives, much less for adjudicating among competing conceptions of collective life or governance.

The point of all this is to suggest that until at least the early twentieth century American political institutions and prevailing theories of politics suited each other rather well, reflecting a largely republican set of political beliefs implemented by largely compatible institutions. In this environment, where representatives were understood to represent territorially-defined interests, most problems with representation could therefore be solved by territorial partitioning of the electorate. For example, one of the most common complaints about inadequate representation during the nineteenth century arose from westward migration: the appearance of newly settled communities gave rise to demands for formal legal recognition and the legislative representation that came with it.\textsuperscript{43} Legislative carving of a new town or county from previously recognized territorial communities thus provided a complete and appropriate solution to the problem: newly-formed communities entitled to legislative representation achieved the recognition they were due, and the new town or county could then become a player in the competitive processes of obtaining central legislative favors and gaining access to centrally controlled resources.

By the late nineteenth and early twentieth centuries, however, the ground began to shift. First, traditionally republican conceptions of

\begin{itemize}
  \item \textsuperscript{40} Michael Zuckerman, Peaceable Kingdoms: New England Towns in the Eighteenth Century 35 (1970).
  \item \textsuperscript{42} See Jean H. Baker, Ambivalent Americans: The Know-Nothing Party in Maryland 94 (1977).
  \item \textsuperscript{43} See Gardner, Representation without Party, supra note 22, at 892.
\end{itemize}
politics began to be displaced by increasingly influential utilitarian theories. Unlike traditional republicanism, which presupposed organic and enduring linkages between place, community, and interest, utilitarianism argued that interests were both individual and highly contingent, and therefore unpredictable. What pleased a person was a matter of personal taste, and a person’s taste was a priori no more likely to be satisfied by one thing than by another. Place and community were thus knocked off their pedestal and demoted to a level of equality with all other potential consumption values. In politics, this meant that the political cleavages that individuals found most salient could just as easily revolve around occupation, social class, ideology, clan, or any of a host of other factors, as around the traditional center of gravity of local community.

By the mid-twentieth century, political scientists had appropriated this new understanding by constructing influential and distinctly anti-republican models of political pluralism. These models rejected a conception of citizenship as revolving around the pursuit of republican virtue and replaced it with one stressing the pursuit of self-interest. By the same token, they also rejected a static conception of politics as jostling among fixed communities in favor of a conception of politics as a fluid, constantly evolving competition among shifting groups organized around contingently salient interests of the moment.

The rise of Progressivism during the late nineteenth and early twentieth centuries launched a second kind of attack on traditional republican institutions. Progressives argued that corruption in politics was pervasive; that the rich and powerful unduly dominated, for their own benefit, the inherited institutions of politics; and that citi-

---

44. For a concrete example of the direct influence of John Stuart Mill’s writings on the members of a late-nineteenth-century state constitutional convention, see Commonwealth v. Wesson, 842 S.W.2d 487 (Ky. 1992).


46. In Bentham’s famous formulation, the only distinction between “push-pin . . . and poetry” is the pleasure people happen to derive from them. Jeremy Bentham, The Rationale of Reward 206 (John & H.L. Hunt 1917) (1825). Accord Bentham, supra note 45, at 43-45.


49. Dahl’s detailed account of “minorities rule” in municipal self-governance probably makes this point as vividly as it has ever been made. Dahl, Who Governs?, supra note 47.
zens, though retaining in theory the capacity to control existing institutions of government for the public good, had permitted themselves to be distracted by objectively irrelevant distinctions of family, neighborhood, and ethnicity.\textsuperscript{50} Progressives consequently mounted a sustained attack on the inherited institutional structure, seeking to replace existing institutions with others they believed more conducive to popular pursuit of rational self-governance in the public interest.\textsuperscript{51} This movement was in many respects extremely successful. Progressives won widespread adoption of measures designed to make voting more rational, such as the secret ballot and the short ballot; to enhance popular control of government, such as primary elections, initiatives, referenda, recall, direct election of U.S. Senators, and female suffrage; and to reduce the role of partisanship in governance, such as the city manager and commission forms of local government, nonpartisanship, and at-large elections.\textsuperscript{52} With the exception of direct election of Senators and the extension of the vote to women, however, all the significant institutional reforms occurred at the state and local levels; the institutions of democratic politics at the national level remained firmly in place.

Lastly, by the mid-twentieth century, popular tolerance for racial exclusions from democratic life waned substantially in most of the country. The civil rights movement focused much of its effort on breaking down racial barriers to voter registration and balloting, achieving a modest initial success in the federal Civil Rights Act of 1957, followed by what would turn out to be a transformational victory in the federal Voting Rights Act of 1965 (VRA).\textsuperscript{53}

All these developments—the eclipse of republicanism, the Progressive reform movement, and the evolution of attitudes concerning race—created serious tensions in the political environment. Existing institutions of democratic politics came increasingly to be seen as out of step with, and inhospitable to, prevailing beliefs about democracy and democratic practice. It was into this frothing cauldron that the Supreme Court finally inserted itself.


\textsuperscript{51} Herbert Croly, The Promise of American Life 315-50 (1909); Buenker, supra note 50, at 188.

\textsuperscript{52} See generally Richard S. Childs, Civic Victories: The Story of an Unfinished Revolution (1952); Benjamin Parke De Witt, The Progressive Movement (Richard T. Ely ed., Macmillan 1915); Hofstadter, supra note 50.

III. THE SUPREME COURT ENTERS THE FIELD

For a long time, the Supreme Court did not concern itself with questions involving the structure or regulation of political practices. For much of American history, opportunities for federal judicial intervention simply did not arise. Throughout the nineteenth century, much of political life was left to private self-regulation. Political parties formed freely, selected candidates by processes of their own choosing, printed their own ballots, and ran campaigns free from governmental oversight.\(^54\) Such election law as existed was almost entirely at the state level, making it a matter for state courts, not federal ones.\(^55\) Indeed, the U.S. Constitution expressly grants to states the authority to regulate federal congressional and presidential elections.\(^56\)

The Supreme Court, moreover, had long taken the position that democratic processes generally, and questions of political representation in particular, were not the business of the federal courts. In a pivotal 1946 ruling, a plurality of the Court ruled malapportionment a nonjusticiable political question, warned against judicial entry into a “political thicket,” and decreed that the only remedy for defects in political representation lay in voluntary legislative action to correct it.\(^57\) The entire panoply of Progressive reforms was implemented, after all, not by judicial intervention, but by legislative action taken in the wake of a highly successful process of political mobilization.

Eventually, however, profound shifts in the social and political environment produced tensions that became too much for the Court to bear, and its resolve to stay out of democratic processes crumbled. As I shall describe shortly, this occurred first in a limited way in *Gomillion v. Lightfoot* (1960),\(^58\) and then more broadly in *Baker v. Carr* (1962),\(^59\) until by the time it decided *Buckley v. Valeo* (1976),\(^60\) the Court was not only heavily involved, but routinely altering the political landscape. This pattern has only continued with decisions such as *Bush v. Gore* (2000),\(^61\) *Citizens United v. FEC* (2010),\(^62\) and,


\(^{55}\) This is still the case today: there is very little election law at the federal level while comprehensive election codes exist in every state. See James A. Gardner & Guy-Uriel Charles, *Election Law in the American Political System* 90 (2012).

\(^{56}\) See U.S. Const. art. I, § 4; id. art. II, § 1.

\(^{57}\) Colegrove v. Green, 328 U.S. 549, 556 (1946).

\(^{58}\) 364 U.S. 339 (1960).

\(^{59}\) 369 U.S. 186 (1962).

\(^{60}\) 424 U.S. 1 (1976) (per curiam).


The Court’s first significant ruling of the modern era in the field of democratic process was *Gomillion v. Lightfoot* (1960), a case that followed closely on the heels, both in time and in subject matter, of perhaps its most important ruling of the twentieth century, *Brown v. Board of Education* (1954), in which the Court ordered an end to racial segregation in public schools. *Gomillion* concerned a law, enacted by the Alabama Legislature at the request of the City of Tuskegee, that altered the boundaries of the city from a perfect square to a meandering 28-sided polygon. After this boundary change, every white resident of Tuskegee still lived within the city, while virtually all of its black residents found themselves outside it.* In a gnostic opinion that laid out more clearly the justices’ horror at this racial gerrymander than their legal reasoning, the Court invalidated the law as an infringement of the right to vote on the basis of race in contravention of the Fifteenth Amendment to the U.S. Constitution.

That *Gomillion* got things started turns out to have been unfortunate in a way; although the case established a template for adjudication that the Court has followed ever since, it is a template that turns out not to be very useful outside the arena of naked racial discrimination. One way in which *Gomillion* got things off to a poor start is that it involves political exclusion in the most literal sense—through the drawing of boundaries to partition, and by partitioning to exclude one portion of, the electorate. Although the case might have been framed as raising questions about the practice of partitioning itself, it was framed instead in a way that assumed the legitimacy of partitioning but treated this particular partition as illicit. As a result, the Court was immediately cast in the role of policing the practice of partitioning the electorate, rather than examining the practice itself on its merits, or inquiring into alternative ways to structure democratic representation that might more directly address the problems of racial exclusion from democratic life.

This was especially unfortunate because the racial gerrymandering undertaken in *Gomillion* represented an extremely unusual form of racial exclusion from democratic participation. Most forms of race-based political exclusion in the United States have not involved the creation of formal geographical boundaries; they have involved in-

---

63. 133 S. Ct. 2612 (2013).
64. 134 S. Ct. 1434 (2014).
68. *See id.* at 345.
instead the application of law, force, and social pressure to exclude members of an existing community from participating in that community’s own democratic practices. They create, in other words, interior boundaries within a society, not reified, external boundaries formally fencing out the disfavored populations.

Second, the gerrymander in Gomillion was undertaken only when more common forms of racial exclusion and suppression had been deemed unpromising. Tuskegee had been since 1881 the home of the Tuskegee Institute (now Tuskegee University), which by the 1950s was an unusually successful, well-regarded, and well-funded black college. During that era, the typical method of first resort in the American South to exclude blacks from participation in a community’s political life was the literacy test. Testing literacy, however, did not recommend itself as a way to exclude black voters associated with the Tuskegee Institute, many of whom held a Ph.D., and were generally far better educated than the city’s white population. The redrawing of the town’s borders to move the Institute and its faculty, staff, and students outside the town was thus an atypical measure of some desperation. Additionally, unlike most other forms of political and social exclusion of blacks, the boundary drawing was not only readily observable, but susceptible essentially to res ipsa loquitur proof of racial animus—no other plausible explanation could account for it.

Finally, the pattern established in Gomillion included deployment of an individual right to solve a problem of democratic practice. The Fifteenth Amendment was available, ready-at-hand, had previously been used by the Court in a few earlier cases, and seemed tailor-made for the kind of problem presented by the Tuskegee racial gerrymander. As a result, the Court gave no thought—and in fairness really did not need to give any thought—to larger, more systemic

69. For innumerable examples, see Branch, supra note 53, and May, supra note 53.
70. For a brief history, see 3 Encyclopedia of African American History 1067-68 (Leslie Alexander & Walter C. Rucker eds., 2010).
73. See Gomillion v. Lightfoot, 364 U.S. 339, 341 (1960) (“If these allegations upon a trial remained uncontradicted or unqualified, the conclusion would be irresistible, tantamount for all practical purposes to a mathematical demonstration, that the legislation is solely concerned with segregating white and colored voters by fencing Negro citizens out of town so as to deprive them of their pre-existing municipal vote.”).
74. These include the “grandfather clause” cases and the “white primary” cases. See Lane v. Wilson, 307 U.S. 268 (1939); Guinn v. United States, 238 U.S. 347 (1915); see also Terry v. Adams, 345 U.S. 461 (1953); Smith v. Allwright, 321 U.S. 649 (1944).
questions concerning appropriate patterns of representation and democratic political life.

Despite the very unusual circumstances of *Gomillion*, the pattern pioneered there was soon applied outside the setting of racial discrimination in a case that went a long way toward cementing it.\(^75\) In *Baker v. Carr* (1962),\(^76\) decided just two years after *Gomillion*, the Court took up the problem of legislative malapportionment, in which legislative election districts contain grossly disparate numbers of voters. In a decision that changed the course of American democracy, the Court reversed its earlier position and held that district population disparities present a justiciable question of constitutional law under the Equal Protection Clause.\(^77\)

Garden-variety malapportionment does not present problems of either race or exclusion. Everyone in a malapportioned district is entitled to vote and to participate fully in politics, and may do so on an equal footing with everyone else in the district.\(^78\) If malapportionment harms processes of political representation, it does so by operation of some defect other than outright exclusion. Upon taking up this problem for the first time, the Court could have dealt with it by deriving or advancing a constitutionally grounded theory of representation. It might have held, for instance, that malapportionment violates some aspect of the way popular sovereignty is meant to work. It could have taken the position, on a kind of pluralist or agency view, that malapportionment erects a barrier to some contemplated degree of government responsiveness to public opinion. Or it might even have invoked a more traditional, republican theory of disenfranchisement of *communities*, understood as territorially defined populations or as territorially defined interests of groups of co-residents.

Instead, the Court approached the problem of malapportionment using the *Gomillion* model. It reached for the most readily available tool—an individual right, the Equal Protection Clause\(^79\)—thereby forcing the case into the mold of an intervention designed to thwart discrimination. But who was discriminating against whom? In the Court’s view, elaborated in later cases, the discrimination effectuated by malapportionment was discrimination in favor of sparsely populated rural areas at the expense of densely populated urban ones.\(^80\)

---

75. In this sense, the phenomenon I am describing may be conceived as a kind of path-dependence. See, e.g., Paul Pierson, *Increasing Returns, Path Dependence, and the Study of Politics*, 94 AM. POL. SCI. REV. 251, 251 (2000).

76. 369 U.S. 186 (1962).

77. *Id.* at 226-37.

78. See *id.* at 335 (Harlan, J., dissenting); Gerken, *supra* note 4, at 506-07.


80. See Reynolds v. Sims, 377 U.S. 533, 568 (1964) (“A citizen, a qualified voter, is no more nor no less so because he lives in the city or on the farm.”). Although the Court never
As in *Gomillion*, the Court again cast itself not as inquiring into the propriety of partitioning of the electorate as a way of organizing political representation, but as policing the practice of partitioning, the legitimacy of which was assumed, and the basis of which was not examined.

Two years later, in the seminal case of *Reynolds v. Sims* (1964), the Court used the new, individual rights lever it identified in *Baker* to effectuate perhaps the most sweeping change in American democratic practice since adoption of the Reconstruction Amendments following the conclusion of the Civil War. In *Reynolds*, the Court for the first time applied the equal protection remedy identified in *Baker* to the apportionment of state legislatures. In a far-reaching decision, the Court held that population disparities among legislative districts violated the right to vote of individuals in overpopulated districts, and that a constitutionally mandated rule of one person, one vote applied to both houses of bicameral state legislatures.

The impact of this decision cannot be overstated. At a stroke, it placed a core feature of a historically decentralized system of representation under central control and destroyed the long-standing structural framework instituting political representation on the basis of place and local community. After *Reynolds*, institutions formerly aimed at achieving representation of communities—of inhabited places—were required to be based instead on shifting, equipopulous groupings of placeless individuals. The Court thus discarded the theory of representation that had long prevailed in the states, not only invalidating it, but deeming it incompatible with what the Court now announced to be constitutionally-grounded notions of equal citizenship.

Although *Reynolds* precipitated wide-ranging and doubtless beneficial changes in the balance of political power across the nation, it established a poor pattern for judicial involvement in the realm of democratic practice. First, the Court did not advance any affirmative theory of political representation; its only theory was negative in said so directly in any of its major malapportionment decisions, the idea of discrimination by rural against urban areas of course has an implicitly racial valence, bringing it squarely within the Court’s emerging comfort zone.

82. *Id.* at 557.
83. See *id.* at 568-72.
84. See Gardner, *One Person, One Vote*, supra note 22, at 1238.
the sense that certain practices—in fact, the old, prevailing practices—were invalid. Second, the Court’s establishment of equal protection as the tool of choice outside the racial context, without a theory of democratic practice and participation to guide its application, was especially problematic. Equal protection outcomes tend to be parasitic on underlying substantive values.\textsuperscript{88} One cannot know whether a person is being treated unequally in any way that counts without first knowing whether that person has a substantive entitlement to the thing he or she has been denied in equal measure, and whether a person has such an entitlement is a question not of equality, but of desert.\textsuperscript{89} By applying equal protection to democratic practices without first specifying the underlying substantive values that are implicated by democratic participation, the Court thus founded its democratic jurisprudence on shifting and unstable ground.

In particular, as Justice Harlan pointed out in dissent, the Court specified neither how much influence citizens should have in a democracy, nor even the outer parameters of what such influence might reasonably be, nor yet any framework within which to think about these questions.\textsuperscript{90} As a result, equal protection outcomes in democracy cases are consistent with a wide array of outcomes that cannot be narrowed except by invoking some antecedent theory of democracy or democratic authority.\textsuperscript{91} Unfortunately, the Court failed to specify what that theory is—it did not indicate, in other words, the proper baseline of comparison for deciding whether democratic influence has been improperly and unequally withheld.\textsuperscript{92} This lacuna has ever since confounded the coherence and utility of constitutional oversight of the political process, and raised many problems that still plague the jurisprudence.

The next two sections focus on two of the most notable problems arising from the template for judicial intervention developed by the Supreme Court in these early cases: its reflexive resort to partitioning of the electorate as a remedy for perceived democratic wrongs or imperfections; and its unthinking deployment of an individual rights model.

IV. Perfectibility Through Partitioning

In cases involving democratic practice in which groups have been excluded or mistreated, the Supreme Court has from the beginning of

\textsuperscript{88} See Peter Westen, \textit{The Empty Idea of Equality}, 95 Harv. L. Rev. 537 (1982).

\textsuperscript{89} See id. at 546-47.


\textsuperscript{92} See Gerken, \textit{supra} note 4, at 507.
its modern jurisprudence displayed a distinct preference for solutions that rely on partitioning the jurisdiction over those that rely on enhancing participation and engagement within the jurisdiction. That is, where members of some group complain that their desire to become full participants in the political life of their community has been thwarted by some officially created obstacle, the Court has preferred not to dwell on ways in which the complaining group might be more fully integrated into existing democratic structures and practices. Instead, it has tended to solve these problems by partitioning the jurisdiction in such a way as to make the complaining minority into a local majority.

An early, important, and in many ways typical example of this approach is White v. Regester (1973). In Regester, black and Chicano populations of two large, metropolitan counties in Texas complained that they had been unconstitutionally excluded from effective participation in the congressional politics of their jurisdictions. The claim involved many moving parts. The plaintiffs complained that they had been victims of a long history of official discrimination in political affairs; that local party processes of choosing candidates were controlled by whites who did not pay sufficient attention to minority communities; that racially divisive campaign tactics had been deployed routinely in white areas; and that, in the case of the Mexican-American plaintiffs, cultural and language barriers and restrictive voter registration practices significantly impeded their political effectiveness. In light of these background conditions, the plaintiffs focused their objections on a specific institutional choice made by the state: its decision to use in these counties large, multimember congressional districts and a place system, in which all candidates ran at-large for specific seats. Ultimately, the plaintiffs argued, the combination of underlying discrimination and the specific institutional choice created conditions in which it was virtually impossible for members of these groups to participate effectively in local congressional politics.

The Court understood these claims perfectly, casting them as challenges to effective participation in democratic processes:

The plaintiffs’ burden is to produce evidence to support findings that the political processes leading to nomination and election were not equally open to participation by the group in question—that its members had less opportunity than did other residents in central office.
the district to participate in the political processes and to elect legislators of their choice.\footnote{88}{Id. at 766.}

Relying on an examination of “the totality of the circumstances,”\footnote{99}{Id. at 769.} the Court held that the plaintiffs had met their evidentiary burden, showing that they had been “effectively excluded” and were “generally not permitted to enter into the political process in a reliable and meaningful manner.”\footnote{100}{Id. at 767.}

When it came to the remedy, however, the Court made a bizarre leap of logic. If the constitutional problem consisted of barriers to participation, then we might expect the remedy to focus on how to lower those barriers and integrate the plaintiff groups effectively into the mainstream of political life in those jurisdictions. Instead, the Court ordered the multimember districts broken up into single-member districts in at least one of which the complaining minority groups would, presumably, become local majorities.\footnote{101}{See id. at 769.} Somehow, the Court suggested, this partitioning of aggrieved minority groups into smaller districts that they could independently control would “bring the community into the full stream of political life of the county and State.”\footnote{102}{Id.} Yet partitioning is the antithesis of overcoming exclusion from participation; it is in fact a different, and in some ways a more extreme, form of exclusion: it takes the complaining group out of the offending jurisdiction and creates a new one in which the group in question no longer has to worry about—much less to engage and work with—the larger group that had previously excluded it.\footnote{103}{See Kathryn Abrams, “Raising Politics Up”: Minority Political Participation and Section 2 of the Voting Rights Act, 63 N.Y.U. L. Rev. 449 (1988). But see Lani Guinier, The Tyranny of the Majority: Fundamental Fairness in Representative Democracy 82 (1994) (arguing that inclusion alone may be futile if it leads to token representation and consistent outvoting, and arguing for alternation in power as an alternative to partitioning).} Certainly, the Court gave no thought to the fact that partitioning the electorate creates new minorities within the newly formed majority-minority districts, or whether the harm of exclusion from participation might now be shifted to such other groups.\footnote{104}{Aleinikoff and Issacharoff call these groups “filler people.” T. Alexander Aleinikoff & Samuel Issacharoff, Race and Redistricting: Drawing Lines after Shaw v. Reno, 92 Mich. L. Rev. 588 (1993). For a well-known example, see United Jewish Orgs. of Williamsburg, Inc. v. Carey, 430 U.S. 144 (1977).}

Since Regester, the solution of partitioning has been applied routinely in many contexts. Where at-large systems have been used for discriminatory purposes, division of the multimember jurisdiction into equipopulous districts has long been the Court’s remedy of choice
in constitutional cases. It is also the Court’s remedy of choice in cases arising under the Voting Rights Act of 1965 (VRA). The VRA, perhaps the most significant American civil rights legislation ever enacted, implements the Fifteenth Amendment’s prohibition of racial discrimination in voting. Section 2(b) of the Act, using language lifted directly from the Supreme Court’s opinion in *White v. Regester*, provides:

A violation . . . of this section is established if, based on the totality of circumstances, it is shown that the political processes leading to nomination or election in the State or political subdivision are not equally open to participation by members of a class of citizens protected by [this statute] in that its members have less opportunity than other members of the electorate to participate in the political process and to elect representatives of their choice.

The statute thus defines an offense in terms of harm to “participation . . . in the political process.” Yet in *Thornburg v. Gingles* (1986), its leading decision on the treatment under the VRA of multimember districts, the Supreme Court ruled that the preferred remedy for denials of participation is not removal of barriers to participation within such districts, but the destruction of multimember districts by partitioning the electorate into single-member districts such that the protected minority is awarded control of one or more of the new districts.

The Court’s habitual resort to partitioning might be perfectly coherent were it led to this remedy by some theory of representation or of democratic process. In fact, however, it has no such theory, and this has produced some significant, persistent problems in the jurisprudence. First, the Court’s lack of an underlying theory leaves unanswered a host of important questions. For example, the act of partitioning the electorate into subgroups necessarily involves decisions about who will have the ability to control an election district, and consequently about who will obtain effective representation in the legislature. Yet without a theory of representation we cannot know who or what is properly represented in a legislature, and thus cannot

---


106. *Id.*

107. *Id.* A Senate Report accompanying the legislation listed numerous factors that might help establish the requisite harm to political participation, including a history of official discrimination in voting; the existence in the jurisdiction of racially polarized voting; the use of electoral procedures that enhance the opportunity for discrimination; denial of access to candidate slating processes; depressed political participation on account of past discrimination in education or employment; racial appeals during campaigns; and the lack of election of minorities to office in the jurisdiction. See *Thornburg v. Gingles*, 478 U.S. 30, 44-45 (1986).

108. 478 U.S. at 50.
make principled decisions about which ways of subdividing the electorate are appropriate and which are not.\(^{109}\)

An example of how these kinds of problems can arise is afforded by *Kramer v. Union Free School District No. 15* (1969).\(^{110}\) That case concerned the validity of a New York law that limited eligibility to vote in school board elections to parents of school-age children and those who owned or rented taxable property in the district.\(^{111}\) The purpose of the law clearly was to confine the school board electorate to those who had some direct or indirect stake in its activities—parents had an interest in the education of their children, and owners and renters of taxable property had an interest in the activities of the school board because it levied school taxes within the district.\(^{112}\) The plaintiff, who lived rent-free in his parents’ house and was thus ineligible under the statute to vote, claimed his right to vote had been infringed.\(^{113}\) The Court agreed,\(^{114}\) but the basis and significance of its ruling is unclear. It did not offer any theory (or if “theory” is too fancy a term, any “account”) of who might be entitled to representation on an elected legislative body, and on what basis. Nor did it suggest that universal suffrage is a constitutional default rule. Instead, the Court “express[ed] no opinion”\(^{115}\) as to whether the state might legitimately restrict the franchise to those who are most directly interested in or affected by governmental actions, but struck down the statute on the ground that it did not advance with sufficient precision the state’s asserted justifications, whether or not they were constitutionally valid.\(^{116}\) The ruling thus clarifies nothing.

Another example is the eye-opening result in *Holt Civic Club v. City of Tuscaloosa* (1978).\(^{117}\) Under Alabama law, cities were permitted to extend the reach of their laws beyond municipal boundaries to unincorporated areas located up to three miles outside city limits.\(^{118}\) Tuscaloosa exercised this option to extend many of its ordinances and regulations to the nearby town of Holt.\(^{119}\) Under the statute, however, Tuscaloosa was not obliged to offer the residents of Holt an opportunity to vote for Tuscaloosa’s city council or mayor.\(^{120}\) As a result,
Holt residents were subject to laws in the making of which they had no voice—a kind of virtual representation soundly repudiated by the American Revolution.  

The Court nevertheless upheld this governance arrangement on the ground that Holt residents did not live in Tuscaloosa, once again failing to buttress its ruling with any account of who is entitled to representation on legislative bodies, and in what circumstances. One might well ask, for example, why residents of Holt, Alabama, were not entitled to representation on a city council that made laws directly binding on them while Mr. Kramer was entitled to representation on a local school board that made laws directly governing the behavior only of others. It may not be necessary for a constitution to provide detailed answers to every question about representation that might arise, but to the extent there is slack in the system, it seems important to know the range of discretion invested in legislatures to define the basis of their own representation.

A second problem arising from the Court’s habitual resort to partitioning despite its lack of a theory of democratic practice or representation is what might be called “partitioning anxiety.” Without any underlying theory of democratically legitimate representation, partitioning is essentially unguided, and the only way corrections to existing arrangements can be made is ad hoc, based on distaste for particular representation schemes. This has led to occasional judicial anxiety about the impact of partitioning.

A clear example of this is the Court’s decisions in a line of cases beginning with Shaw v. Reno (1993). As explained above, the VRA, as construed by the Court, in some circumstances requires states to partition the electorate so as to give blacks and other protected populations a substantial degree of control over an appropriate number of election districts. States got the message, and proceeded to comply. Having set this pattern in motion, however, the Court soon began to exhibit a case of severe anxiety.


122. Holt Civic Club, 439 U.S. at 70.

123. See James A. Gardner, How to Do Things with Boundaries: Redistricting and the Construction of Politics, 11 ELECTION L.J. 399, 417-18 (2012) (arguing that constitutionalized norms of democracy can be adequately regulated by defining a range of permissible choices, requiring jurisdictions to make explicit choices about institutional arrangements, and then exercising judicial enforcement by requiring those jurisdictions to act consistently with their publicly declared commitments).


Shaw v. Reno concerned the constitutionality of a tortuous district drawn by North Carolina for the purpose of sweeping up black populations in different parts of the state in sufficient numbers to give them political control of a congressional district. The state drew the district for no purpose other than to comply with the VRA. Yet the Court found the shape of the district too “bizarre,” and the state’s reliance on race as the operative criterion for partitioning voters too single-minded, causing it to invalidate the district as racially discriminatory under the Equal Protection Clause. In so doing, the Court placed states in the impossible position of attempting to walk a fine line, the location of which remained obscure: on one hand, a state violated the VRA if it did not try hard enough to provide racial minorities with districts they could control; on the other, a state violated the Equal Protection Clause if it tried too hard.

All this is a direct and predictable consequence of the Court’s insistence that the electorate be partitioned coupled with its refusal to supply any guidance as to what kind of representation a partitioned electorate ought to enjoy. Having demanded the creation of majority-minority districts, the Court balked at the implications of its own requirement, and its instruction to states amounted more or less to the following directive: partition, but not like this. The only thing the Court has done since then to make things easier for states engaged in redistricting is its recent invalidation of a significant provision of the VRA in Shelby County v. Holder (2013). Following that ruling, states need be much less fearful of liability under the VRA, leaving them to fear realistically only liability under the Equal Protection Clause, a more difficult kind of case for plaintiffs to prove up. Nevertheless, the ruling does nothing to clarify how states ought to draw election districts.

But perhaps the most serious problem arising from the Court’s heavy reliance on partitioning to solve problems of democratic pro-

128. Id. at 634.
129. Id. at 644, 655-56.
130. See id. at 642-45.
132. See 133 S. Ct. 2612, 2631 (2013).
133. Proof of a violation of the Equal Protection Clause requires proof of intentional discrimination, Washington v. Davis, 426 U.S. 229 (1976), whereas proof of a violation of the VRA requires at most, in a section 2 case, a showing of disparate impact, without regard to intent. Thornburg v. Gingles, 478 U.S. 30, 74 (1986). Under Section 5, prior to Shelby County, changes by a covered state to its electoral laws were presumed discriminatory until proven by the state to be racially benign or neutral. See Georgia v. Ashcroft, 539 U.S. 461 (2003).
cess is that it leads political actors to focus not on the fairness or content of political processes within a district, but on acquisition of tactical control over the boundaries of districts.\textsuperscript{134} In the world the Court has helped to create, democracy is constructed not only by processes of voice and mutual engagement within a jurisdiction, but also by manipulating who is in and who is out of that jurisdiction. The route to success in politics thus often lies less in offering a set of normative commitments attractive enough to appeal to voters than in sending one’s opponents into exile by partitioning them out of the territory.

This problem plagues American democracy. It manifests itself most often in persistently contentious processes of redistricting in which political actors contest for power unguided by transparent and binding legal principles.\textsuperscript{135} Lacking constitutionally grounded standards they are required to respect, redistricting authorities typically fall back on coarse imperatives of power and partisanship. Redistricting thus is treated not as an occasion to bring democratic practice into conformity with democratic ideals—which remain unspecified—but as an opportunity to cement temporary partisan advantage into place for the next ten years until a new census is taken and the process repeats itself.

V. \textbf{AVAILABILITY: THE UNGUIDED DEPLOYMENT OF GENERIC INDIVIDUAL RIGHTS}

As explained above, the Court’s reluctance to develop a constitutionally-grounded account of the norms that structure and guide American democratic processes created a vacuum at the heart of its jurisprudence of democratic practice, a vacuum that, inevitably, had to be filled with something. The previous section demonstrated how the Court filled this vacuum in part by falling back on habitual forms of problem-solving by reflexively deploying partitioning of the electorate as the standard treatment for a host of democratic ills. This part describes another way in which the Court filled the vacuum left by its refusal to specify a constitutional theory of democratic practice: by reaching for the handiest and most readily available tool—though not necessarily the most appropriate one—to resolve constitutional challenges to the democratic legal order. That tool was individual rights.\textsuperscript{136}

\textsuperscript{134} The title of a recent book by a prominent political scientist says it all. \textit{See} Charles S. Bullock III, \textit{Redistricting: The Most Political Activity in America} (2010).

\textsuperscript{135} States and even electorates (in initiative states) have tried to shape the process by providing some limited forms of normative guidance, but without much success. \textit{See} Gardner, \textit{Representation without Party}, supra note 22, at 894-98.

\textsuperscript{136} On the disadvantages of a rights-based approach compared to one based on constitutional structure in cases addressing the constitutionality of democratic practice and process, \textit{see} Samuel Issacharoff & Richard H. Pildes, \textit{Politics as Markets: Partisan Lockups of
A. The Reign of Equal Protection

For about two decades, the Court’s main tool for resolving disputes over democratic practices and processes was the Equal Protection Clause. Following Baker v. Carr and subsequent one person, one vote cases, the Court routinely turned to equal protection in dozens of cases involving challenges to franchise restrictions,\(^{137}\) malapportionment,\(^{138}\) restrictions on ballot access,\(^{139}\) regulation of political parties,\(^{140}\) and many other issues. Yet outside of cases involving obvious racial discrimination, the Equal Protection Clause was not well suited to carry the burden of the Court’s reliance.

The Equal Protection Clause is clearly useful in cases challenging official racial discrimination in democratic processes because redressing racial discrimination is what the Clause was principally designed to achieve. Moreover, the fact that principles of equality are generally parasitic on underlying substantive norms—their application, in other words, depends upon the existence of an independently supplied normative baseline\(^ {141}\)—does not pose a problem in cases of racial discrimination for the obvious reason that the Constitution itself clearly and emphatically establishes such a normative baseline: purposeful racial discrimination is not to be tolerated in any official endeavor.\(^ {142}\)

The Equal Protection Clause, however, is much less useful for resolving problems of democracy that do not present claims of racial discrimination. Its limited suitability to resolving such claims is perhaps most clearly revealed by the awkward gyrations the Court was forced to undergo simply to find the Clause applicable to the most basic controversies involving voting. In these cases, what the Court wanted was a constitutional right to vote. On its face, this presented


\(^{140}\) See Buckley v. Valeo, 424 U.S. 1, 93-107 (1976) (per curiam).

\(^{141}\) See Westen, supra note 88. See also supra notes 87-91 and accompanying text.

a problem: the U.S. Constitution does not expressly grant in any provision a right to vote in federal elections, and indeed it incorporates as the criterion of eligibility to vote in federal elections whatever standards states have chosen to adopt for eligibility to vote in their own legislative elections. Thus, as the Court has said on more than one occasion, “the Constitution . . . does not confer the right of suffrage upon any one . . .”

Still, a constitution may be found to confer rights by means other than express enumeration. The Court might, for example, have inferred the existence of a right to vote from the structure and purpose of the Constitution’s many provisions establishing representative democracy. Doing so, though, would presumably have forced the Court to acknowledge that the Constitution implicitly establishes some principles of democratic self-rule, something it has not wished to do. To avoid doing so, the Court chose instead to find the right to vote buried awkwardly in the Equal Protection Clause. It consequently ruled that although the U.S. Constitution does not oblige a state to allow anyone in particular to vote, once a state chooses to extend the franchise to anyone at all, the Equal Protection Clause requires that individuals be permitted to participate in elections “on an equal basis with other qualified voters whenever the State has adopted an elective process for determining who will represent any segment of the State’s population.” Thus, bizarrely, the right to vote came to be lodged in a provision that does not speak of voting;

146. This is the premise of the Ninth Amendment. See U.S. Const. amend. IX.
147. In a very early case, the Court appeared as though it might move in this direction. In Wesberry v. Sanders, 376 U.S. 1 (1964), the Court invalidated malapportionment of congressional districts on the basis of inferences drawn from a structural provision, Article I, § 2 of the U.S. Constitution, which requires that members of the House be elected “by the People of the several States.” However, it immediately retreated from this turn to structure in its next case, Reynolds v. Sims, 377 U.S. 533 (1964), in which it invalidated malapportionment in state legislative districts on equal protection grounds. The only other significant decisions in the field that the Court has reached on structural grounds are U.S. Term Limits, Inc. v. Thornton, 514 U.S. 779 (1995), in which the Court relied on the Qualifications Clauses to invalidate state-imposed term limits on members of Congress; and Cook v. Graham, 531 U.S. 510 (2001), in which the Court invalidated a state law requiring disparaging statements to be placed on the ballot next to the names of candidates who refused actively to pursue a federal constitutional amendment to impose term limits on members of Congress. In the latter ruling, the Court relied on implicit structural conceptions of proper representation and legislative judgment. Graham, 531 U.S. at 510.
does not confer on its own a right to vote; and cannot itself supply a normative decision principle for resolving disputes about the proper extent of the right to vote.

This latter problem, at least, could have been solved were the Court willing to supply a principle of decision by extracting from the Constitution some theory about what the vote is for and why, and in what circumstances, citizens are entitled to have it. As we have seen, however, this is precisely what the Court declines to do. As a result, equal protection analysis in the area of voting rights becomes unmoored and haphazard as the Court searches for, or lurches between, principles adequate to resolve its cases.

In no subfield of election law has this problem more thoroughly crippled the Court’s decision making capacity than in the field of redistricting, an area that presents perhaps the most pressing problems in all of American democratic practice. When redistricting raises issues of racial discrimination, the Court’s tools for dealing with it are more than adequate.149 Difficulties arise in handling a much more widespread problem, the problem of partisan gerrymandering, in which redistricting is performed so as to provide one party with an outsized and undeserved advantage over its opponents.150 The Court’s attempt to handle this problem by resort to principles of equal protection has been a spectacular failure.

On three occasions in the last thirty years the Court has tried and failed to identify a constitutional standard under the Equal Protection Clause for adjudicating the constitutionality of partisan gerrymandering.151 Its failure is directly traceable to the Court’s deployment of equal protection without an underlying theory to identify a baseline of proper representation, departure from which can therefore be understood as illicit gerrymandering.152 Justice Kennedy, who cast the deciding vote in Vieth v. Jubelirer (2004),153 admitted this frankly in his opinion:


150. To the extent that members of the Court have been able even to settle on a definition of the problem, they have defined it, for example, as conferring on the dominant party a degree of power that is “excessive,” League of United Latin Am. Citizens v. Perry, 548 U.S. 399, 463 (2006) (Stevens, J., concurring in part and dissenting in part), or “too much” in relation to its fair or appropriate share, Vieth v. Jubelirer, 541 U.S. 267, 344 (2004) (Souter, J., dissenting), or as the “continued frustration of the will of a majority of the voters or effective denial to a minority of voters of a fair chance to influence the political process,” Davis v. Bandemer, 478 U.S. 109, 133 (1986) (White, J., for a four-justice plurality).

151. See League of United Latin Am. Citizens, 548 U.S. at 404; Vieth, 541 U.S. at 305; Davis, 478 U.S. at 110.

152. Gerken, supra note 4, at 506-07; Gerken, supra note 7, at 414, 420.

153. 541 U.S. 267.
Because there are yet no agreed upon substantive principles of fairness in districting, we have no basis on which to define clear, manageable, and politically neutral standards for measuring the particular burden a given partisan classification imposes on representational rights. Suitable standards for measuring this burden, however, are critical to our intervention.154

He went on to issue an earnest appeal for help in identifying an appropriate baseline:

That no such standard has emerged in this case should not be taken to prove that none will emerge in the future. Where important rights are involved, the impossibility of full analytical satisfaction is reason to err on the side of caution. . . . This possibility suggests that in another case a standard might emerge that suitably demonstrates how an apportionment’s de facto incorporation of partisan classifications burdens rights of fair and effective representation.155

This is as far as the Court has ever gotten in a partisan gerrymandering case, and consequently legislatures engaged in the task of redistricting have little reason to fear effective judicial enforcement of any constitutional prohibition on gerrymandering.

B. The Empire of the First Amendment

By the mid-1970s, the Court began to find that the Fourteenth and Fifteenth Amendments no longer reliably supplied decision rules for every kind of case dealing with democratic practice and process that the Court was willing to accept. Consequently, if the Court was to continue to adjudicate such cases by deploying off-the-shelf, readily available individual rights, it would have to import some other right into the democratic arena. In *Buckley v. Valeo* (1976),156 the Court turned decisively to the First Amendment guarantee of freedom of speech. In *Buckley*, the Court invoked the First Amendment to determine the constitutional validity of the Federal Election Campaign Act, the most comprehensive piece of federal campaign finance regulation ever enacted, invalidating numerous portions of the Act in large part on the ground that they unduly impaired constitutionally protected speech.157 Not long after, in *Anderson v. Celebrezze* (1983),158 a case challenging state rules restricting access of independent presidential candidates to the election ballot, the Court took the significant step of repudiating the Equal Protection Clause as its main workhorse in ballot access cases. Instead, the Court announced

154. *Id.* at 307-08.
155. *Id.* at 311-12.
156. 424 U.S. 1, 51 (1976) (per curiam).
157. *Id.* at 19-22, 50, 54, 58.
158. 460 U.S. 780, 786-87 n.7 (1983).
without explanation that it would thenceforth analyze ballot access restrictions under the First Amendment right of freedom of association, a second-order right derived by implication from the freedom of speech.\footnote{\textit{NAACP v. Alabama}, 357 U.S. 449, 462 (1958).} The Court went on in \textit{Anderson} to invalidate the restriction at issue on the ground that it burdened constitutionally protected association between candidates and their supporters.\footnote{\textit{Id.} at 792-95, 806.}

In subsequent cases, the Court has invoked the First Amendment to adjudicate nearly every kind of dispute involving regulation of the democratic process. It has deployed the First Amendment not only in cases revolving around campaign speech, campaign finance, and ballot access, but also in cases dealing with restrictions on voting,\footnote{See \textit{Burdick v. Takushi}, 504 U.S. 428, 441-42 (1992).} political parties,\footnote{See \textit{Eu v. S.F. Cnty. Democratic Cent. Comm.}, 489 U.S. 214, 233 (1989).} primary elections,\footnote{See \textit{Clingman v. Beaver}, 544 U.S. 581, 593 (2005); Cal. Democratic Party v. Jones, 530 U.S. 567, 586 (2000); Tashjian v. Republican Party of Conn., 479 U.S. 208, 214 (1986).} and election integrity.\footnote{See \textit{Burson v. Freeman}, 504 U.S. 191, 211 (1992).} Some justices have suggested that even partisan gerrymandering cases would be more tractable if handled under the First Amendment.\footnote{See \textit{Vieth v. Jubelirer}, 541 U.S. 267, 314-15 (2004) (Kennedy, J., concurring); \textit{id.} at 324 (Stevens, J., dissenting).} In fact, so versatile has the Court found the First Amendment that it has begun to approach its democracy cases as though the First Amendment is the \textit{only} provision in the entire Constitution of the slightest relevance to the system of representative democracy it institutionalizes. This odd approach might be harmless if the Court’s resort to the First Amendment represented merely some kind of well-understood judicial synecdoche, in which the First Amendment is invoked as a kind of short-hand reference to the entirety of the constitutional scheme. Unfortunately, that is not the case. The Court’s understanding has become close to literal; the First Amendment has become for the Court essentially a one-provision constitution, complete in itself, capable of solving any and every problem of democracy for which judicial review may be had.

This approach has been costly, and the main casualty has been the First Amendment, which has in these cases been stretched beyond all recognition. Although it is indisputably handy, dangling tantalizingly at the top of the Bill of Rights like a fly before a trout, freedom of speech simply is not an instrument well-suited to the work of
adjudicating many of the complex questions that arise concerning democratic practice and procedure.  

The most notable example of the inadequacy of the right to free speech to handle problems for which it has been deployed is the area of campaign finance. In a series of cases beginning with *Buckley v. Valeo* (1976) and continuing through the Court’s recent decisions in *Citizens United v. FEC* (2010) and *McCutcheon v. FEC* (2014), the Court has deployed the freedom of speech to decide the constitutionality of laws that restrict the giving and spending of money in connection with election campaigns for public office. In so doing, the Court has afforded the same degree of First Amendment protection to giving and spending money in election campaigns as it does to campaign speech itself; in the Court’s jurisprudence, there is no constitutionally significant difference between political spending and political speaking.

The Court’s indiscriminate use of the First Amendment has been harshly criticized for decades, and there is no need to rehearse that criticism here. Suffice it to say that money is tied to speech only loosely, and that equating the regulation of money spent to buy speech with regulation of the speech itself proves far too much and thus bites far too deeply into democratically legitimate and justifiable regulatory regimes. Furthermore, a crucially important component of the First Amendment doctrine that the Court imported into the arena of democratic practice is a long-standing judicial tradition of very nearly absolute opposition to the regulation of fully protected forms of speech. As a result, the Court’s importation into the democracy arena of a pure free speech regime, unmodified to suit the context, has led, predictably, to shockingly deregulatory results.

---

166. See sources cited supra note 136, arguing for the superiority of a structural approach, or at least an approach to rights that places them in the service of structural considerations.

167. 424 U.S. 1, 51 (1976) (per curiam).


171. A small but suggestive collection of such critiques may be found in GARDNER & CHARLES, supra note 55, at 684-91.

172. This objection was immediately raised in *Buckley* itself by Justice White. See *Buckley*, 424 U.S. at 262-64 (White, J., concurring in part and dissenting in part).

Thus, in *Buckley*, the Court gutted the Federal Election Campaign Act by invalidating nearly every limitation on campaign spending contained in the Act.174 It continued to invalidate regulatory limitations on political spending in a long series of cases,175 and in some instances invalidated limitations on campaign contributions to candidates as well.176 In its 2010 decision in *Citizens United*, the Court shocked observers by invalidating a century-old prohibition on direct political spending by corporations.177

Critical to the Court’s ruling in these cases is its rejection in *Buckley* of the basic legitimacy of one of Congress’s principal reasons for enacting restrictions on campaign spending: to redress inequality in political influence between the rich and the poor,178 an interest the Court deemed “wholly foreign to the First Amendment.”179 As a matter of run-of-the-mill First Amendment free speech doctrine, government attempts to orchestrate a fair balance of views expressed in everyday discourse in civil society might well be viewed with extreme skepticism.180 Speech made in the course of democratic processes intended to constitute a binding expression of the popular will, however, is no ordinary speech,181 and the Court’s importation into this arena of an existing, off-the-shelf First Amendment regime seems effectively to have blinded the Court to extremely significant differences in context—differences of goals, stakes, complexity, and countervailing values.182

---


181. *See, e.g.*, DENNIS F. THOMPSON, JUST ELECTIONS: CREATING A FAIR ELECTORAL PROCESS IN THE UNITED STATES 115 (2002) (“[E]lections and the campaigns leading up to them may be considered more a part of government than a part of politics that influences government. The standards that control the conduct of elections should therefore be determined more by collective decision than by individual choice.”). To similar effect, see Dennis F. Thompson, *Election Time: Normative Implications of Temporal Properties of the Electoral Process in the United States*, 98 AM. POL. SCI. REV. 51, 60-62 (2004); Saul Zipkin, The Election Period and Regulation of the Democratic Process, 18 WM. & MARY BILL RTS. J. 533, 575-76 (2010).

182. For an overview, see GARDNER & CHARLES, supra note 55, at 670-78. A comprehensive account of the best reasons for and against campaign finance restrictions may be found in JACOB ROWBOTTOM, DEMOCRACY DISTORTED: WEALTH, INFLUENCE AND DEMOCRATIC POLITICS (2010). An alternative and more nuanced approach is that of the Supreme Court of Canada, which has found the constitutional protection of democracy to embody numerous distinct values which are capable in many circumstances of coming into conflict. *See* Dawood, *supra* note 6, at 252-57.
Nor is First Amendment doctrine particularly well-suited to deal with other issues to which the Court has applied it. Consider, for example, the Court’s handling of questions of ballot access, which arise when laws regulate the conditions under which candidates may have their names placed on the election ballot. For three decades, the Court has adjudicated such cases under the First Amendment right of association, evaluating ballot access rules in terms of the degree to which they burden association between candidates and their followers.\textsuperscript{183} There are many ways to think about ballot access. One might plausibly say that what is at stake in ballot access cases is presenting voters with an appropriate and meaningful range of choices,\textsuperscript{184} or that governmental restrictions on ballot access raise issues of incumbent self-protection or partisan self-dealing,\textsuperscript{185} or that ballot access restrictions potentially limit the optimal degree of political competition.\textsuperscript{186} But to proceed as though the only constitutionally relevant question concerns the ability of candidates and their supporters to associate is downright strange. Enjoyment of association with others may be a worthwhile benefit of group political participation, but it is not the main goal, nor does the printing of a candidate’s name on the ballot in any meaningful way enhance the quantity or quality of association between the candidate and his or her supporters.\textsuperscript{187} A more plausible explanation for why the Court analyzes ballot access in these terms, then, is its desire to make use of a readily available, off-the-shelf right, in lieu of thinking about new or alternative frameworks to apply in such cases.

VI. CONCLUSION

The phrase “constitutional jurisprudence” generally conjures up the image of a court working hard to develop a versatile and internally coherent body of doctrine that furnishes appealing solutions to pressing legal problems while bringing constitutional text and purpose into harmonious alignment with judicial implementation. If so, then one is hard-pressed to apply the term to the incoherent and haphazard body of law developed by the U.S. Supreme Court to adjudicate problems of democratic practice and process. To the extent the


\textsuperscript{185} See Issacharoff & Pildes, supra note 136, at 683-87.

\textsuperscript{186} See id.

\textsuperscript{187} For a skeptical view of this proposition, see Tashjian v. Republican Party of Conn., 479 U.S. 208, 235 (1986) (Scalia, J., dissenting).
Court can be said to have such a jurisprudence at all, it is a largely accidental one that the Court has stumbled into through habit, the vagaries of doctrinal availability, and a susceptibility to path-dependent decision making, rather than one that has been deliberately crafted through the application of judicial imagination and diligence.

What is required, clearly, is for the Court to set aside its squeamishness about “political theory” and do in the area of democratic politics precisely what it has done in other areas of constitutional structure, such as federalism and the horizontal separation of powers: develop a theory of what the Constitution is trying to do and how it strives to go about it. There is no reason why the Court cannot derive from the Constitution’s structural provisions, underlying principles, and historic democratic commitments an account of the nature and appropriate processes of representative democracy. That would be a useful first step in a much-needed program to undo the damage caused by the Court’s failure to provide a sensible foundation for judicial review in this singularly important area of constitutional law.
COPYRIGHT'S MERCANTILIST TURN

GLYNN S. LUNNEY, JR.

ABSTRACT

Over the last twenty years, arguments for broader copyright have taken an increasingly mercantilist turn. Unable to establish that broader copyright will lead to more or better original works, as the Constitution and the traditional economic framework require, proponents have begun arguing for broader copyright on the basis of revenue and jobs. Rampant unauthorized copying is theft or piracy, proponents insist, depriving copyright owners of revenue and destroying jobs. Whether or not it leads to more or better works, broader copyright will increase revenue to copyright owners and thus increase employment in the copyright industries. This increased employment, on its own, justifies broader copyright, or so proponents contend. In this Article, I critically reexamine this argument and show that it is empty.

I. INTRODUCTION

Over the last twenty years, justifications for broader copyright protection have taken an increasingly mercantilist turn. In the recent debates over the Protect Intellectual Property Act (PIPA) and the Stop Online Piracy Act (SOPA), proponents did not seriously argue that these measures would enhance welfare by encouraging the production of more and better works of authorship. Rather, they argued that these bills would increase revenues to domestic copyright owners and thereby create jobs. This shift from neoclassical welfare economics to mercantilist justifications for policy is not unique to PIPA.

1. See infra notes 24-31 and accompanying text.

2. The term “mercantilist” or “mercantilism” has been applied to several different political or economic approaches. Originally, it referred to approaches that sought to maintain a favorable balance of trade in order to bring gold and silver into an economy while removing it from other nations’ economies. Over time, however, this emphasis on the money supply diminished, and mercantilism came to refer to a collusive relationship between government and mercantile interests where, in exchange for paying levies and taxes to support the nation-state, government would enact policies to protect business interests against competition. See ROBERT B. EKELUND, JR. & ROBERT D. TOLLISON, MERCANTILISM AS A RENT-SEEKING SOCIETY: ECONOMIC REGULATION IN HISTORICAL PERSPECTIVE (1981); ROBERT B. EKELUND, JR. & ROBERT D. TOLLISON, POLITICIZED ECONOMIES: MONARCHY, MONOPOLY, AND MERCANTILISM (1997); ELISE S. BREEZIS, MERCANTILISM, in THE OXFORD ENCYCLOPEDIA OF ECONOMIC HISTORY 484 (2003) (“[Adam] Smith’s second criticism of mer-
and SOPA, however. Rather, it has become a defining feature of United States trade policy with respect to copyright and intellectual property, more generally, over the last few decades. Moving away from the tenets of free trade, trade policy in the intellectual property arena has sought increasingly to protect domestic industries from foreign competition and to ensure thereby more revenue for and more jobs in those industries within the United States.

This switch from neoclassical welfare economics to mercantilism is, in one sense, entirely understandable. To justify broadening copyright under neoclassical welfare economics, proponents would need to establish three propositions: first, for any given work of authorship, broader copyright would lead to more revenue for that work than would narrower copyright; second, the prospect of such increased revenue would lead individuals to devote additional resources to, and hence create, more or better works of authorship; and third, society would value such more or better works more highly than the alternative uses to which the resources would otherwise have been devoted. While the first proposition might seem self-evident, the second and third certainly are not and as it turns out, establishing any of these propositions, let alone all three, has proven extremely difficult.

With respect to the second proposition, for example, for hundreds of years copyright has relied on the simple, yet fundamental, premise that more revenue for copyright owners means more and better original works of authorship. Yet, the relationship between revenue and creative output is not so straightforward. Consider the music industry. As various popular press accounts have repeatedly touted, record sales have fallen sharply since Napster started the file-sharing ball rolling. In 1999, record sales, whether in physical or electronic format, stood at $20.4 billion (in constant 2012 dollars). In 2012, such sales amounted to just over $7 billion. The record industry attributes this decline to unauthorized consumer copying, file sharing, or in their preferred parlance, “piracy” or “content theft.” Empirical studies dispute this. Nevertheless, while the role unauthorized copying...
has played in this decline remains unclear, the decline itself cannot be disputed.

Yet, regardless of whatever may have caused this decline, we have not seen a corresponding fall in the production of new music. Three empirical studies have examined music output since Napster opened its doors in 1999, and all three found that music output has continued to grow.\(^7\) In my own study, I use regression analysis to account for other factors, such as declining costs, and show that the decline in revenue was associated with an increase in high quality music output.\(^8\) Reduced revenue was associated with fewer new artists, as copyright's traditional economic account suggests. However, the same reduction in revenue was also associated with increased output from existing artists. As I predicted in 2001,\(^9\) it appears that file sharing has reduced the excess incentives that our existing copyright laws would have otherwise provided for our most popular artists and authors. Facing reduced revenue from any given song, our most popular artists have chosen to create and disseminate more songs in order to reach their desired income and standard of living. Moreover, because the increased output from existing artists exceeded the decreased output from having fewer new artists, the reduction in revenue that has occurred in the music industry since file sharing began has been

---


\(^8\) Lunney, supra note 7.

associated with a net increase in the production of new hit songs, *ceteris paribus*.

Contrary to popular expectations, then, file sharing seems to have had a positive net impact on the production of original works of music, and it seems to have had that effect precisely because it reduced revenue to copyright owners. Given that the Constitution and Supreme Court opinions plainly state that copyright’s purpose is not to maximize copyright owners’ revenue, but “[t]o promote the Progress of Science,”¹⁰ there would seem to be no rational basis for measures, such as SOPA and PIPA, designed to reverse the digital revolution and curtail file sharing and other forms of consumer copying.

Recognizing the difficulty and, perhaps the futility, of attempting to justify such measures under the traditional economic and constitutional framework, proponents of broader copyright have sought to reframe the debate. Rather than attempt to establish that broader copyright enhances welfare or increases the production of copyright-ed works, proponents of broader copyright have argued instead that broader copyright will generate more revenue for copyright owners and, in turn, more jobs in the copyright industries. While this alternative argument has some superficial appeal, in large part, that appeal arises because the jobs argument nicely complements the traditional “more and better works” justification. With broader copyright, not only can we increase creative output, but we can also create more jobs and ensure the proverbial starving artist a livable wage—all with one stroke of our legislative pen. Despite this superficial attractiveness, the real question is whether the jobs argument can justify broader copyright on its own, independent of copyright’s traditional justification. In other words, even if broader copyright meant no increase in creative output—or as is more likely today given copyright’s extreme breadth, a measurable decrease in creative output—can we nevertheless justify broader copyright on the grounds that it will create more jobs in the copyright industries?

We cannot. This attempt to refocus the debate on revenue and jobs represents a classic mercantilist ploy and, as Adam Smith established more than three hundred years ago, is seriously flawed.¹¹ While creating jobs seems like a good and desirable role for govern-

---

¹⁰ U.S. CONST. art. I, § 8, cl. 8; see also Feist Publ’ns, Inc. v Rural Tel. Serv. Co., 499 U.S. 340, 349 (1991) (“The primary objective of copyright is not to reward the labor of authors, but '[t]o promote the Progress of Science and useful Arts.’ ” (quoting U.S. CONST. art. I, § 8, cl. 8)); Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) (“The immediate effect of our copyright law is to secure a fair return for an 'author's' creative labor. But the ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good.”); United States v. Paramount Pictures, Inc., 334 U.S. 131, 158 (1948) (“The copyright law, like the patent statutes, makes reward to the owner a secondary consideration.”).

ment, particularly in today’s economy, governments do not, in the end, create jobs. Needs, wants, and desires do. Through regulations, taxes and spending, and other government action, governments can shift jobs from one sector of the economy to another, by encouraging the satisfaction of some desires, while discouraging the satisfaction of others. As a general rule, however, over the long run, creating jobs in one sector, through government action, necessarily entails losing jobs elsewhere.\textsuperscript{12}

Consider the argument that broader copyright creates jobs. For broader copyright to create jobs in the copyright industries, it must increase revenue to copyright owners. Yet, for the copyright industries to receive more revenue, consumers must pay more for works of authorship. Broader copyright, after all, does not generate revenue from thin air. It has to come from somewhere. If consumers have to pay more for works of authorship, they will have less to spend on everything else. Thus, more revenue for the copyright industries necessarily means less revenue for other sectors of the economy. If more revenue for copyrighted works means more jobs for the copyright industries, presumably less revenue everywhere else means fewer jobs elsewhere in the economy.

In that sense, the jobs argument for SOPA and PIPA or for broader copyright, more generally, represents a perfect example of Frederic Bastiat’s Broken Window Fallacy.\textsuperscript{13} Writing in 1850, Bastiat emphasized the need to account in economics and politics both for that which is seen and that which is not seen. (“Ce qu’on voit et ce qu’on ne voit pas.”) When a boy breaks a shopkeeper’s window, the shopkeeper must employ a glazier to fix it. If we focus solely on the employment of the glazier—that which is seen—then one might conclude that the government should hire children to go around breaking windows in order to increase the employment of glaziers in the economy. Howev-

\textsuperscript{12. As Professor Machlup has explained in the patent context:

\begin{quote}
It is easy to conceive of the possibility that such allocation [of productive resources to invention] is too meager. But can there ever be too much? Is not more research and development always better than less? Is it possible that too much is devoted to the inventive effort of the Nation? This depends on what it is that is curtailed when inventive activity is expanded. More of one thing must mean less of another, and the question is, what it is of which there will be less. . . . Whenever permanent economic policies—not just war or depression measures—are discussed, sound economics must start from the principle that no activity can be promoted without encroaching on some other activity. More of one service or product must mean less of another.
\end{quote}

\textsuperscript{13. Frederic Bastiat, That Which Is Seen and that Which Is Not Seen, in Essays on Political Economy 72 (David A. Wells trans., N.Y., G. P. Putnam’s Sons 1877) (1850).}
er, as Bastiat cautioned, we must also account for that which is not seen. Because the shopkeeper had to spend his money on the glazier, he could not spend that money elsewhere: on new shoes or a new book for his library. When we account for this lost spending elsewhere—that which is not seen—we find that the broken window generates no net stimulus to employment. The glazier earns more; but whoever would have received that money but for the broken window—whether cobbler, bookseller, or another—earns exactly that much less.

The mercantilist argument for broader copyright suffers from much the same fallacy. It urges us to focus solely on that which is seen—the increased revenue and enhanced employment broader copyright brings to the copyright industries. It asks us to ignore that which is not seen—the reduced revenue and diminished employment broader copyright brings to every other sector of the economy. Once we account for both that which is seen and that which is not seen, we find the mercantilist argument for broader copyright entirely empty. Just as the broken window generates no net stimulus for the economy, so too does broader copyright. Whatever increased revenue broader copyright generates for the copyright industries, it simply takes from elsewhere in the economy.

Of course, arguing for broader copyright in order to increase the employment of authors and artists is not precisely the same as arguing for a government policy urging children to break windows in order to increase the employment of glaziers. Breaking windows destroys something of value; encouraging authorship creates something of value—more or better works of authorship. Yet, this difference, while important, does not diminish the relevance of Bastiat’s parable. Even though authorship is a productive activity, we still must account for that which is not seen. But for broader copyright, there would have been more revenue for and hence more employment in other sectors of the economy, and presumably that alternative employment would have been productive as well. Although the alternative employment would not create more or better works of authorship, it would create something else of value to society. The question thus becomes whether we would value the additional works of authorship that we assume broader copyright will bring forth more

---

14. While the Broken Window parable is the first and perhaps most well-known in Bastiat’s essay, he actually presents a number of examples and, as we shall see, addresses this sort of trade-off directly, particularly in his section entitled “Theatres and Fine Arts.” See infra text accompanying notes 101-11.
than or less than the something else that we would have had but for broadening copyright.\textsuperscript{15}

By so accounting for that which is not seen, we find that the mercantilist jobs mantra is not a shortcut; it does not and cannot avoid the hard question of valuing alternative uses of the available resources. To justify broader copyright on the grounds that it will create more jobs, proponents still must show that those copyright jobs produce something that is more valuable to society than the goods or services that the alternative jobs would have produced. While phrased slightly differently, this is essentially the same proposition proponents of broader copyright have been unable to establish under the traditional economic and constitutional framework.

Demonstrating both a loss of revenue to copyright owners and a loss of associated jobs in the copyright sectors of the economy is not, therefore, sufficient to justify broader copyright, whether in the form of PIPA and SOPA or otherwise. Labeling the behavior at issue as “piracy” or “content theft” cannot provide the necessary justification either. The only way to justify further government intervention is for the proponents of broader copyright to demonstrate that the widespread consumer copying and distribution that the Internet has made possible has: (1) led to a reduction in the expected revenue associated with any given work; (2) that this reduction in expected revenue has led to a reduced output of works of authorship; and (3) that the lost output in copyrighted works would have been more valuable than the alternative output elsewhere in the economy. However, proponents of broader copyright have neither made this showing nor does it appear that they can. Arguing that broader copyright increases revenue to the copyright industries and thereby leads to more employment in those industries proves nothing all.

This article explores these issues in turn. Section II begins with a review of the arguments proponents advanced in support of PIPA and SOPA and notes the emphasis on lost revenue and jobs. In Section III, we examine the neoclassical welfare approach to defining the optimal scope of copyright. As part of this discussion, we take a look at how digital technology has changed the music industry and examine in detail the effects consumer copying has had on creative output. In Section IV, we turn to a critique of the mercantilist approach to copyright. Taking Bastiat’s lessons to heart, Section IV demonstrates that revenue and jobs alone are not a sufficient justification for government intervention absent associated net welfare losses. It also demonstrates that pursuing a copyright policy that attempts to max-

\textsuperscript{15} As Professor Baxter has explained: “This is the classic economic criterion for optimal allocation.” William F. Baxter, Legal Restrictions on Exploitation of the Patent Monopoly: An Economic Analysis, 76 Yale L.J. 267, 268 n.7 (1966).
imize revenue for and employment by the copyright industries inevitably leads to welfare losses.

II. ILLUSTRATING THE MERCANTILIST TURN

On January 18, 2012, democracy happened in the United States. Until that point, it appeared that SOPA and PIPA were headed for passage. Following the lead of the Senate Judiciary Committee, which had passed PIPA in May 2011, the Chairman of the House Judiciary Committee, Representative Lamar Smith (R-Tex), introduced SOPA on October 26, 2011, with an initial group of twelve bipartisan co-sponsors. According to its sponsors, the bill was intended to shut down foreign websites that made available to United States consumers unauthorized copies of copyrighted works (described as “rogue” websites, in keeping with the view that labels alone suffice to make plain wrongdoing). It did so both directly, by providing for court orders that would require Internet service providers to block access to such websites (the so-called DNS blocking remedy) and indirectly by providing for court orders that would bar: (i) online advertisers and payment facilitators from conducting business with such sites; and (ii) search engines from listing such sites. The House Judiciary Committee held an initial hearing on November 16, 2011, featuring four witnesses who supported the bill and one, Katherine Oyama, Google’s Copyright Counsel, who opposed it.

Despite proponents’ attempts to limit testimony on the issues, opposition to and concerns about the bill began to mount. As a result,

16. For example, in the second day of mark-up hearings on SOPA, Representative Issa (Calif.), an opponent of SOPA, acknowledged that opponents of SOPA were going to lose. Markup of H.R. 3261, Stop Online Piracy Act: Hearing Before the H. Comm. on the Judiciary, 112th Cong. 28 (2011) [hereinafter SOPA Markup] (statement of Rep. Issa), available at http://judiciary.house.gov/_files/hearings/pdf/transcript12162011.pdf (“Mr. Chairman, it is very clear, we are going to lose here today. No, let me rephrase that, we are going to lose eventually, and we are going to lose in the worst possible way. We are going to lose without all the facts, we are going to lose without the process being open in the way that I would hope it will be in the new year.”).


18. Id. § 102(c)(2)(B)–(D).

19. See SOPA Markup supra note 16, at 82 (statement of Rep. Sensenbrenner), available at http://judiciary.house.gov/_files/hearings/pdf/transcript12152011.pdf (“I am concerned that when we had the hearing on this bill as far as the DNSSEC provisions were concerned, none of the six witnesses were able to address this issue.”); id. at 88 (statement of Rep. Chaffetz) (“I understand that there is a problem, but I worry that this is the wrong remedy. I was trying to think of a way to try to describe my concerns with this bill, but basically we are going to do surgery on the Internet, and we haven’t had a doctor in the room tell us how we [sic] going to change these organs. We are basically going to reconfigure the Internet and how it is going to work without bringing in the nerds, without bringing in the doctors.”); id. at 95, 98 (statement of Rep. Lungren) (“One of my problems with this bill is we have not had the benefit of technical experts to appear and testify before us, and certainly on the same platform. . . . When we had that last hearing, there wasn’t a
Representative Smith introduced an amended version of the bill on December 15, 2011. Mark-up hearings were then held for two days before Congress adjourned. At the end of the December 16th mark-up session, Representative Smith stated that he expected “to resume markup at the next earliest practicable day that Congress is in session.” By January 16, 2012, the number of cosponsors for SOPA had swelled from the initial twelve to thirty-one, and Representative Smith had scheduled the next mark-up session for January 20th.

However, just two days before the scheduled mark-up, on January 18, 2012, the English Wikipedia, Reddit, and an estimated 7,000 other smaller websites coordinated a service blackout in order to raise awareness on SOPA and PIPA. Together with petition drives and other protests, this online activism generated a firestorm like nothing Congress had ever seen before—at least on copyright issues. In response, on January 20, 2012, Representative Smith cancelled the scheduled mark-up hearing and postponed further consideration of SOPA.

My focus for now is not on the Internet as a means of overcoming the collective action problems that have long plagued the opponents of broader copyright: it is on the arguments proponents of SOPA

single person who could answer the technical questions, and they all admitted that, even though a couple of them still opined.”).

20. Id. at 9-11.
21. Id. at 57.
22. See H.R. 3261.
24. As I have explained elsewhere:

In dealing with the Copyright Act, we should bear in mind that it directly benefits a well-organized special interest group, authors and publishers, at the expense of a more dispersed group, the public. Given a statute [sic] with such a distribution of benefits and burdens [sic], public choice theory predicts that over time the statute will inevitably come to favor more and more the desires of the special interest group at the expense of the more dispersed group. When combined with some superficially plausible rationale that can serve to screen the legislator’s motivations, the concentrated group’s disproportionate ability to raise money that can be used—whether in the form of campaign contributions, bribes, or for expert opinions that back the group’s position—to convince legislators to favor the concentrated group’s position, has proven unfortunately [sic] persuasive in convincing our elected representatives to serve the special interest at the expense of the general public. That our elected representatives have therefore broadened copyright’s protection to include these additional rights provides no assurance that there is adequate [sic] justification for these rights. Awarding authors these rights will often be the result of interest group pressure, combined with a superficially plausible explanation to cover the legislators’ actions.
and PIPA used to justify their proposals. From the outset, proponents of SOPA and PIPA said little about what one might consider the traditional justifications for broader copyright. In fact, almost nothing at all was said about how these “rogue” websites, if not addressed, would lead to fewer works of authorship. Certainly, no attempt was made to show that not only would fewer works be forthcoming but that there would be fewer works than would be socially optimal. Rather, proponents of SOPA and PIPA emphasized the need for legislation to protect American jobs.

For example, in joining with Representative Smith to introduce SOPA, Representative Bob Goodlatte (R-Va) issued a press release, entitled Goodlatte Introduces Legislation to Protect American Jobs, explaining his support for SOPA. Given its title, the basis for his support was not surprising:

Piracy denies individuals who have invested in the creation and production of these goods a return on their investment thus reducing the incentive to invest in innovative products and new creative works. The end result is the loss of American jobs. Estimates indicate that IP theft costs the U.S. economy over $100 billion a year and results in the loss of thousands of American jobs.25

Similarly, during the November 16th hearing, Michael O’Leary of the Motion Picture Association of America sought to justify enactment of SOPA as follows:

Fundamentally, this is about jobs. The motion picture and television industry supports more than two million American jobs in all 50 states. The 20 states and Puerto Rico represented by this Committee are home to 1.7 million American jobs supported by the motion picture and television industry, including more than 525,000 direct motion picture and television industry jobs. About 12 percent of those are directly employed in motion picture and television production and distribution, jobs paying an average annual salary of nearly $79,000. Those are not just the people whose names you see on the marquee in front of the theater – they’re the hardworking people behind the scenes, from the carpenter who built the set, to the costumer and make-up artist who helped bring each character to life, to the Foley artist who created the sound effects.26


26. The “Stop Online Piracy Act”: Hearing on H.R. 3261 Before the H. Comm. on the Judiciary, 112th Cong. 72 (2011) (statement of Michael P. O’Leary, Senior Exec. Vice President, Global Policy and External Affairs, on behalf of the Motion Picture Association of
Following the November 16th hearing, another of SOPA’s co-sponsors, Representative John Conyers, Jr. (D-Mich.), reiterated the jobs theme in expressing his support for SOPA:

I have always stood by artists, and it is for this reason that I want to see the Stop Online Piracy Act become law. The bill is of vital importance to protecting American jobs and artisans, protecting the American consumers from dangerous counterfeits, and ensuring the very vitality of American culture.

[...]

We can protect and promote American jobs, perhaps millions of them, by getting a bill like this to President Obama’s desk for his signature as soon as possible.  

Even when forced to postpone further consideration of SOPA, Representative Smith reiterated his support for congressional action in this area, on the basis, yet again, of jobs: “American intellectual property industries provide 19 million high-paying jobs and account for more than 60 percent of U.S. exports. The theft of America’s intellectual property costs the U.S. economy more than $100 billion annually and results in the loss of thousands of American jobs.”

This emphasis on jobs as a justification for expanded copyright protection is not simply a result of our recent recession. In 2007, before either the recession or SOPA and PIPA had arrived on the scene, the copyright industries had already begun using jobs as an argument to support copyright expansion. In that year, the Institute for Policy Innovation released a study purporting to show that for the year 2005, piracy accounted for losses to the U.S. economy of $58 billion in output, over 370,000 jobs, and $2.6 billion in tax revenue.

The emphasis on jobs has also not been restricted to copyright. In March 2012, the U.S. Patent and Trademark Office, together with the Economics and Statistics Administration, issued its report on the economic impact of intellectual property in the United States. Of

---


28. Statement from Chairman Smith, supra note 23.


the report’s ten principal findings, six related directly to the relationship between intellectual property and jobs, including:

1. A finding that IP-intensive industries account directly for 27.1 million American jobs, or 18.8 percent of all employment in the economy, in 2010;

2. A finding that IP-intensive industries account indirectly for another 12.9 million supply chain jobs in the United States, and thus account directly or indirectly for 27.7 percent of all jobs in the economy;

3. Findings with respect to job creation and job growth in the IP-intensive industries; and

4. Findings that jobs in IP-intensive industries pay more highly than jobs in non-IP-intensive industries.\(^\text{31}\)

Moreover, it’s quite clear that this emphasis on jobs as a justification for broader intellectual property will not end with the defeat of PIPA and SOPA. In July 2012, the Chamber of Commerce issued its cry for action on counterfeiting and piracy by reiterating the relationship between intellectual property and jobs: “IP-intensive industries directly and indirectly support more than 55 million American jobs – jobs that pay 30% higher wages than those in other industries – and account for $5.8 trillion in national output.”\(^\text{32}\)

Now to some extent, we can readily reject many of these jobs-based arguments for broader copyright on the relatively simple ground that the job loss numbers are as bogus as a three-dollar bill. As the GAO explained in a recent report,\(^\text{33}\) reliable estimates of both the revenue losses and the job losses caused by unauthorized copying are nearly impossible to generate for two reasons. First, the amount of unauthorized copying itself is difficult to estimate.\(^\text{34}\) Second, the rate at which unauthorized copying substitutes for authorized purchases is also difficult to estimate.\(^\text{35}\)

\(^{31}\) Id. at vi-viii.


\(^{33}\) U.S. Gov’t Accountability Office, GAO-10-423, INTELLECTUAL PROPERTY: OBSERVATIONS ON EFFORTS TO QUANTIFY THE ECONOMIC EFFECTS OF COUNTERFEIT AND PIRATED GOODS 15-16 (2010) (“Most experts we spoke with and the literature we reviewed observed that despite significant efforts, it is difficult, if not impossible, to quantify the net effect of counterfeiting and piracy on the economy as a whole.”).

\(^{34}\) Id. at 16.

\(^{35}\) Id. at 17.
Yet, the purpose of this paper is not to critique the job loss numbers that proponents of PIPA and SOPA have offered. Others have admirably tackled that topic already.36 The question that I would like to address is whether the creation of additional jobs in the copyright industries, standing on its own, can ever provide a legitimate, alternative basis for copyright expansion. If we knew for certain that broadening copyright in a certain way would create some thousands of jobs and have no effects other than job creation, would those jobs provide a basis for legislative action? To begin to answer that question, the next section begins with a review of copyright’s traditional welfare justification, both as a matter of theory and empirically through an examination of the impact file sharing has had on the music industry.

III. COPYRIGHT’S TRADITIONAL JUSTIFICATION: THE FEAR OF THE COPYING COMPETITOR AND ITS LIMITS

For more than four hundred years, copyright’s central justification has remained essentially unchanged. In the absence of copyright protection, we fear that competitors will quickly copy new works of authorship, undercut the price for authorized copies and thereby reduce the incentive authors will receive for creating and disseminating original works. Indeed, in a perfectly competitive world, competitors would copy new original works immediately and by offering their copies at a price set to cover the competitors’ marginal costs, would eliminate any economic incentive for creating and disseminating an original work. In such a world, in the absence of copyright, we might not have too few original works but none at all. For more than four hundred years, this fear that the incentives for authorship might prove insufficient in the absence of copyright has provided copyright’s central justification. The

Stationer’s Guild used it to justify its monopoly over printing as long ago as 1586.\textsuperscript{37} And it remains today, remarkably unchanged.\textsuperscript{38}

In the same way, the central limit on copyright has remained equally unchanged. Copyright raises the price of books and other copyrighted works.\textsuperscript{39} That higher price simultaneously provides the incentive to create original works and limits access to existing works. Given the higher price, some consumers will no longer be able to afford authorized copies of an original work. The search for optimal copyright is therefore thought to entail a search for the optimal balance between incentives and access. As Lord Thomas Macaulay expressed in his speech to the House of Commons in 1841:

It is good that authors should be remunerated; and the least exceptionable way of remunerating them is by a monopoly. Yet monopoly is an evil. For the sake of the good we must submit to the evil; but the evil ought not to last a day longer than is necessary for the purpose of securing the good.\textsuperscript{40}

Or as Professor Landes and Judge Posner more recently offered:

Copyright protection—the right of the copyright’s owner to prevent others from making copies—trades off the costs of limiting access to a work against the benefits of providing incentives to create the

\textsuperscript{37} The Stationer’s Guild argued:

And further if privileges [that is, copyright] be revoked no books at all should be printed, within short time, for commonly the first printer is at charge for the Author’s pains, and some other such like extraordinary cost, where an other that will print it after him, comes to the Copy gratis, and so may he sell better cheaper than the first printer, and then the first printer shall never utter [that is, sell] his books.


\textsuperscript{38} Almost exactly four centuries after the Stationer’s Guild articulated it to the Star Chamber, Professor Landes and Judge Posner used it in their attempt to justify copyright. They argued:

In [the] absence [of copyright protection] anyone can buy a copy of the book when it first appears and make and sell copies of it. The market price of the book will eventually be bid down to the marginal cost of copying, with the unfortunate result that the book probably will not be produced in the first place, because the author and publisher will not be able to recover their costs of creating the work.


\textsuperscript{39} I am not particularly interested in whether we characterize these higher prices as “monopoly” or not. See Edmund W. Kitch, \textit{Elementary and Persistent Errors in the Economic Analysis of Intellectual Property}, 53 VAND. L. REV. 1727, 1729-38 (2000).

\textsuperscript{40} THOMAS BABINGTON MACAULAY, \textit{A Speech Delivered in the House of Commons on the 5th of February, 1841, in 8 THE WORKS OF LORD MACAULAY 195, 199 (Lady Trevelyan ed., 1897), available at http://yarchive.net/macaulay/copyright.html.}
work in the first place. Striking the correct balance between access and incentives is the central problem in copyright law.\footnote{Landes & Posner, \emph{supra} note 38, at 326; see Computer Assocs. Int'l, Inc. v. Altai, Inc., 982 F.2d 693, 696 (2d Cir. 1992) ("[C]opyright law seeks to establish a delicate equilibrium. On the one hand, it affords protection to authors as an incentive to create, and, on the other, it must appropriately limit the extent of that protection so as to avoid the effects of monopolistic stagnation."); Niva Elkin-Koren, \emph{Copyright Policy and the Limits of Freedom of Contract}, 12 \textit{Berkeley Tech. L.J.} 93, 100 (1997) ("[C]opyright law seeks to balance the level of incentives to create and the interest in maximizing access to information once created. Finding the correct balance between access and incentives is the central task of copyright policy."); Robert A. Kreiss, \emph{Accessibility and Commercialization in Copyright Theory}, 43 \textit{UCLA L. Rev.} 1, 4 (1995) ("To function properly, copyright law must strike a balance between the rights given to copyright authors and the access given to copyright users."). \textit{But see} William M. Landes & Richard A. Posner, \emph{The Economic Structure of Intellectual Property Law} 11 (2003) (acknowledging that while the balance between incentives and access is important, it cannot account for everything).}

In devising an optimal copyright system, this supposed balance between incentives and access has become the central guide: too little copyright and we will have too few original works; too much and we will not be able to enjoy the works we have. Only through the appropriate balancing of incentives and access will we have neither too little nor too much copyright develop—copyright will be just right, or at least, that’s the conventional wisdom.

Yet, there are problems with this approach. First, as it is usually phrased, the attempt to balance incentives and access is incomplete and mischaracterizes the issue. Second, even if we knew what to balance, we may not have the information that we need to devise a copyright system that strikes the balance appropriately.\footnote{See, e.g., Joseph P. Liu, \emph{Owning Digital Copies: Copyright Law and the Incidents of Copy Ownership}, 42 \textit{WM. \\& MARY L. Rev.} 1245, 1310 (2001) ("Although the minimalist framework, as a theoretical matter, accurately captures the essential considerations underlying copyright law (that is, the basic balance between access and incentives), lack of information and numerous practical difficulties in applying the framework serve to limit its ability to generate clear or effective results.").} Part A reviews what we know about devising an optimal copyright system, given the information available. Part B then presents some new empirical data examining the impact of the Internet and digital technology on the music industry in the United States to see what light, if any, that new information may cast on the question of optimal copyright. Here we find, somewhat surprisingly, that while revenue is down, output in the music industry is up.

\section*{A. Neither Too Much nor Too Little, the Search for “Goldilocks” Copyright}

Even if we think of an exclusive right to prohibit copying as a form of property, property rights in information and other goods characterized by non-rivalrous consumption, so-called “public” goods, lack the intrinsic desirability of property rights in goods characterized by ri-
valorous consumption, so-called “private” goods. While copyright may be necessary to encourage the creation and release of an original work of authorship, once the work has been created and released, copyright becomes simply an encumbrance. By raising the price of access to original works, it creates a deadweight loss for those consumers unable to afford the higher price.

But the deadweight loss from higher prices is not the only cost that copyright creates. Copyright prohibits a wide range of potential uses of copyrighted works absent the copyright owner’s consent. This effectively imposes a licensing requirement for those uses and thereby creates transaction costs that would not otherwise exist. Where those transaction costs exceed the potential gains from trade for particular uses, it blocks the use directly and thus creates externalities. And by requiring the copyright owner’s permission, it limits our ability to put an original work to all of its highest and best uses. Instead of a work being put to all of its valuable uses—as it should be given its public good character—it is put only to those uses of which the copyright owner approves.

As a result, an original work of authorship, once created and disseminated, will be less valuable to society with copyright than without, and it will be less valuable with more copyright than with less.

Moreover, copyright generates a second cost, one that does not arise from limiting access to existing works. Specifically, copyright can push popular authors and artists onto the backward-bending portion of the labor supply curve and thus perversely lead to fewer works from the most popular authors and artists. The labor supply curve becomes backward-bending at the point where a worker’s earnings become high enough that the worker begins to value leisure more highly than labor. If you pay me one thousand dollars per hour, I will work more than if you pay me one hundred dollars per hour. But if you pay me ten thousand dollars per hour or one hundred thousand dollars per hour, at some point I will begin to work less. When my earnings become high enough, I will want to spend my time doing the things that I most enjoy, rather than working. As an exam-

43. See Glynn S. Lunney, Jr., Copyright, Derivative Works, and the Economics of Complements, 12 Vand. J. Ent. & Tech. L. 779, 802-03 (2010).

44. As Arrow has explained: “It is necessary to distinguish between the realized social benefit and the potential social benefit, . . . which, in this case, means the sale of the product at postinvention cost, c’. Clearly, the potential social benefit always exceeds the realized social benefit.” Kenneth J. Arrow, Economic Welfare and the Allocation of Resources for Invention, in NAT'L BUREAU OF ECON. RESEARCH, THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 622 (1962); see also William W. Fisher III, Reconstructing the Fair Use Doctrine, 101 Harv. L. Rev. 1659, 1715-17 (1988); Landes & Posner, supra note 38, at 341-42 (explaining that consumer and producer surplus would be higher with less copyright protection).

45. See Lunney, supra note 9, at 890-92.
ple of this, Mike Scherer has traced the productivity of Giuseppi Verdi, one of the most influential 19th century Italian composers, following the introduction of copyright in northern Italy. After copyright was introduced and Verdi’s per-opera earnings increased, Verdi substantially reduced the number of operas he wrote and released each decade.46 My own favorite example of this is country-and-western singer, Garth Brooks. After releasing his first album in 1989, Mr. Brooks became one of the top-selling solo artists of all time. Yet, rather than encourage Mr. Brooks to write and perform more music, copyright ensured that Mr. Brooks earned so much from his music that he could afford to and did retire in 2001, at the age of 39. While he announced his return to music in 2009 and is doing stage shows in Las Vegas, since his retirement, Garth Brooks has not been the creative musical force that he was during the 1990s.

Rather than the unhelpful catch-phrase, “access,” copyright’s principal costs within the markets for the creation and dissemination of copyrighted works arise from its uniform scope of protection and the resulting overprotection of works that would have been created and disseminated with less or no copyright. When we consider the range of original works that could potentially be created, we find that some works will be brought forth with no or very little copyright protection, while others require more copyright protection. Yet, copyright remains a fundamentally one-size-fits-all system. If we increase copyright protection in order to bring forth those works that require more protection, we will also provide that same increased protection to all of the preexisting works.47 Because of these works’ public-good character, overprotecting the preexisting works reduces their social value,

46. F.M. Scherer, The Emergence of Musical Copyright in Europe from 1709 to 1850, 5 REV. ECON. RES. ON COPYRIGHT ISSUES 3, 11 (2008) [hereinafter Scherer, The Emergence of Musical Copyright]; see also F.M. Scherer, QUARTER_NOTES AND BANK NOTES: THE ECONOMICS OF MUSIC COMPOSITION IN THE EIGHTEENTH AND NINETEENTH CENTURIES 194 (2004) (explaining that new copyright policies made a “substantial difference” to Verdi’s compositional effort); F.M. Scherer, The Evolution of Music Markets, in 1 HANDBOOK OF THE ECONOMICS OF ART AND CULTURE 123, 138 (Victor A. Ginsburgh & David Throsby eds., 2006) (offering Giuseppi Verdi as another example of copyright pushing an artist onto the backward-bending portion of the labor supply curve). As Professor Scherer explained:

Verdi’s abundant surviving correspondence makes it clear that the maestro viewed opera composition as exhausting drudgery. As his wealth accumulated, Verdi reduced his compositional effort — from 14 operas in the 1840s to seven in the 1850s, two in the 1860s, and one each in the succeeding three decades. The reduction in effort cannot be attributed to declining ability; some of Verdi’s great operas are among the handful of late compositions. Rather, his correspondence makes clear, the higher “price” elicited for each opera made it possible to reduce effort along a classic backward-bending supply curve.

Scherer, The Emergence of Musical Copyright, supra note 46, at 11.

47. By preexisting, I mean that the works would have been created and disseminated with less or no copyright protection.
and it is this overprotection cost for the preexisting works that we must balance against the social value added by any additional works that broader copyright brings forth. Specifically, overprotection reduces the social value of works that would have been created and disseminated with less or no copyright; and it ensures that the most popular authors and artists receive payment for their work that so far exceeds their reservation cost for that work that it likely reduces the output of these authors and artists. Against these overprotection costs, broader copyright’s principal benefit is the value of those additional works that would not be forthcoming but for broader copyright.

This shift in emphasis, from balancing benefits from additional works against costs for preexisting works, rather than benefits against costs for a particular work, may seem trivial, but it is critically important. When we consider the potential range of original works that broader copyright could bring forth, we will likely find, to the extent works of authorship exhibit similar cost and demand structures, that there is some normal distribution in the number of works that increasing copyright protection will bring forth, as Figure 1 illustrates.

Figure 1. Distribution of Additional Work Brought Forth as Copyright Protection Increases.

The idea in Figure 1 is that different works require different levels of copyright protection to ensure their expected profitability and

48. In addition to these costs, copyright may also skew the incentives for authors, encouraging authors to create popular works at the expense of great works. See Lunney, supra note 9, at 888-90.
hence creation. Thus, copyright protection at level $A$ ensures the expected profitability of works at $A$ and to the left of $A$. Given the copyright protection available, works to the right of $A$ remain unprofitable and will not be forthcoming. If we want to increase output in the copyright sector, we need to increase the level of copyright protection, moving, for example, from $A$ to $B$ on Figure 1. Such an increase in copyright protection will ensure the expected profitability of a larger category of works and will bring forth those additional works that lie between $A$ and $B$.

Yet, if we expand protection from level $A$ to level $B$, copyright’s uniform term and scope of protection means that we will provide the broader protection not only to the additional works that the increased copyright protection brings forth, but also to the preexisting works that would have been brought forth with less or no copyright protection. Moving from $A$ to $B$ will thus increase the overprotection costs associated with these preexisting works. Nevertheless, as Figure 1 suggests, in moving from $A$ to $B$, the overprotection costs associated with preexisting works will not prove overwhelmingly large. For such a move, the preexisting works consist only of those works that lie under the distribution curve to the left of $A$. This is only a relative few, both in absolute number and relative to the number of additional works the shift to $B$ brings forth.

On the other hand, if we move from a lot of copyright protection to even more, moving from $C$ to $D$ on Figure 1, the overprotection costs become substantial. For this move, the preexisting works consist of all of those that lie under the distribution curve to the left of $C$. This is a very large number of preexisting works, both in absolute terms and relative to the number of additional works that the shift to $D$ brings forth. Even if one believes that an efficient licensing market will help reduce the overprotection costs for any given work, when we add up the overprotection costs for such a very large number of preexisting works, those costs likely become prohibitive.

It’s true that, even when framed properly, this balance between the costs and benefits of broader copyright remains an empirical question. Nevertheless, it does not take detailed economic information to get an accurate, if very rough, sense for the likely costs and benefits for some copyright and copyright-like expansions. Consider the 1998 copyright term extension, which extended copyright from life-plus-fifty years to life-plus-seventy years. On the benefit side, the first question should have been: How many additional works will an additional twenty years of protection bring forth? The probable answer, given the uncertainty and discounting associated with any rev-
enue so far in the future, is statistically indistinguishable from zero.\textsuperscript{49} On the other side of the balance, adding twenty years to copyright’s term has a considerable social cost. Specifically, it ties up the vast number of works that would have been created even without the term extension—essentially all works—for an additional twenty years. In short, the proposed extension generated all costs and no benefits.\textsuperscript{50} Or consider the recent proposals for fashion design protection. On the benefit side, the initial question again is: How many additional designs will we get by providing fashion design protection? My own sense, here, given the strong reputational rents already available for creative designers,\textsuperscript{51} is very little. On the other side of the balance, the question is: What is the cost to society of protecting all of the fashion designs that we would have gotten even without such protection? Given the great variety of designs being produced without such protection today, my own sense, again, is that the likely costs of overprotecting all of the preexisting designs would be substantial. As with copyright term extension, the costs of fashion-design protection will likely prove so high, and the benefits, if they exist at all, would be so small, it would seem that the only reasonable answer we can reach is that fashion-design protection would reduce social welfare.\textsuperscript{52}

While this balancing of costs and benefits can therefore prove useful, it remains incomplete. So far in balancing the costs and benefits


\textsuperscript{50} Lord Macaulay made the same basic point in arguing against copyright term extension in England in the 19th century. As he explained in a speech to the House of Commons in 1841:

\textbf{MACAULAY, supra note 40, at 198-202.}

\textsuperscript{51} For a formal model of reputation rents and innovation, see Glynn S. Lunney, Jr., Patent Law, the Federal Circuit, and the Supreme Court: A Quiet Revolution, 11 SUP. CT. ECON. REV. 1, 58-63 (2004).

of broadening copyright, we have relied on two usually unspoken assumptions. First, we have focused solely on the markets for copyrighted works and have ignored how broader copyright might interact with imperfections in other markets. Such a partial equilibrium approach necessarily assumes, inter alia, that all other markets are complete (i.e. that there are no externalities in other markets) and perfectly competitive.53 Second, we have also assumed that consumers will be willing to substitute unauthorized copies of an original work for authorized copies,54 even if that threatens the creation of a given work.

Neither assumption matches the real world, however. Other markets suffer from imperfections too.55 Free riders, positive externalities, and copying are present not only in the market for original works of authorship but are ubiquitous features of virtually every market. Similarly, while consumers may sometimes substitute an unauthorized copy of an original work for an authorized copy, they will not always do so. Particularly when that substitution begins to threaten the incentives necessary to ensure an original work’s creation, we should expect a consumer’s own self-interest to lead the consumer to purchase the authorized copy or otherwise contribute to the creation of the original work.56 Acknowledging this more complicated reality and adopting more realistic assumptions on these issues sharply narrows copyright’s optimal scope.

As I have shown elsewhere, when we acknowledge that other markets face free riding, positive externalities, and copying, copyright will tend to ensure the allocation of available resources to their highest and best use when it ensures that those who invest in the creation and dissemination of original works, more generally, experience a return neither greater than nor less than the return available for creative work in other fields.57 Such an approach leads to the op-

53. See Lunney, supra note 24, at 488 n.13; see also JOAN ROBINSON, THE ECONOMICS OF IMPERFECT COMPETITION 317-18 (1933).

54. Perhaps, the assumption is that consumers cannot tell the difference between authorized and unauthorized copying, but that can be addressed by a law requiring proper labeling.

55. See, e.g., ROBINSON, supra note 53, at 51, 88-89; Glynn S. Lunney, Jr., Copyright’s Price Discrimination Panacea, 21 HARV. J.L. & TECH. 387, 406-09 (2008) (“In short, markets for creativity, wherever they are found in the economy, are neither complete nor perfectly competitive.”).

56. See Glynn S. Lunney, Jr., Copyright, Private Copying, and Discrete Public Goods, 12 TULANE J. TECH. & INTELL. PROP. 1, 17, 20, 32 (2009).

57. See Lunney, supra note 24, at 599-601; see also Lunney, supra note 55, at 447 (showing that such a rough equivalence promotes an efficient allocation of resources in a second-best world). Terry Fisher has acknowledged the importance of this point but has expressed the concern that my proposed solution “would sacrifice most of the economic benefits highlighted by Demsetz and Goldstein.” William Fisher, Theories of Intellectual Property, in NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY 168, 182 (Ste-
timal allocation of available resources in two respects. First, it would lead additional resources to be allocated to creating additional works if that represents the highest valued use of those resources. But, second, it would not lead additional resources to be allocated to creating additional works if there is some other use of those resources that would create more value. As it turns out, this standard also has the advantage that it is relatively easy to implement given the empirical information available. For example, if we are trying to determine copyright’s optimal duration, determining whether the costs outweigh the benefits for extending copyright from life-plus-fifty years to life-plus-seventy years was relatively easy, even under the partial equilibrium approach. However, attempting to use that approach to define whether the optimal copyright term is one year, five years, ten years, or something longer is much more difficult.

In contrast, if our goal is to ensure authors and copyright owners a period of exclusivity approximately the same as the lead-time available to other innovators, the empirical data to determine the optimal

phen Munzer ed., 2001). The concern seems to be that without a legal right to control certain users, creators of original works will not receive appropriate pricing signals. As Paul Goldstein has argued: “The logic of property rights dictates their extension into every corner in which people derive enjoyment and value from literary and artistic works. To stop short of these ends would deprive producers of the signals of consumer preference that trigger and direct their investments.” PAUL GOLDSTEIN, COPYRIGHT’S HIGHWAY 178-79 (1994). This is utter nonsense. It fundamentally misunderstands the nature of pricing in competitive markets. As an example, consider Professor Goldstein’s argument that we would have too few books suitable for making into movies unless an author has a right to charge a separate licensing fee for making a movie from the book. See GOLDSTEIN, supra, at 178-81. Using his reasoning, we could equally well argue that we will have too few off-road capable vehicles unless carmakers have the right to charge a separate licensing fee for off-road use. Without a property right to price such use separately, how will car-makers receive an appropriate price signal for off-road capable vehicles? As we all readily recognize, the off-road use argument is wrong, but for the same reasons, so is Goldstein’s argument for a derivative work right. In competitive markets, prices are a function of costs, not of consumer values. As a result, if vehicles suitable for both on- and off-road use cost more to bring to market than vehicles suitable for only on-road use, then so long as there is sufficient demand, the price for off-road capable vehicles will be higher. If they are not more costly, then the price will not be higher. We do not need an “off-road use” right that a carmaker can separately license in order to generate appropriate pricing signals. It’s the same for books. If a book suitable for movie use is more costly to write, then the market price for such books will be higher. If not, then not. Again, no “movie making” derivative work right is necessary for appropriate pricing signals. I have explained elsewhere the differences between cars and books that might justify a derivate work right. See Lunney, supra note 24, at 628-30, 632, 641. In short, a derivate work right can be justified only if it is necessary to ensure the creation of the book in the first place or if there is only going to be one movie-version of the book. As it turns out, both rationales are persuasive only if the associated feature film is likely to be a natural monopoly, even in the absence of copyright. See Lunney, supra note 43, at 814. While movies have historically had such a character, digital technology is reducing that natural monopoly character, as well as the natural monopoly character of radio airplay. As it does so, rights that were sensible given the associated natural monopolies, such as the derivative work right and the public performance right, will likely become increasingly unnecessary and, indeed, undesirable.
copyright term is readily available.\textsuperscript{58} For innovations protected by trade secrets, innovators on average have a lead-time of two to twelve years.\textsuperscript{59} Given patentee’s decisions on whether to pay the maintenance fees or not, patents, on average, last eleven years.\textsuperscript{60} If copyright competes for resources, such as creativity, with these other creative, but not copyrightable sectors of the economy, then to avoid either an underproduction or an overproduction of original works, copyright’s original fourteen-year term seems pretty close to ideal.\textsuperscript{61} Such an approach also suggests that copyright’s principal, if not exclusive, focus should be on mechanical duplication, as that is the principal difference between copying the creativity in original works and copying creativity elsewhere in the economy.\textsuperscript{62}

\textsuperscript{58} Some may argue that, even given the proper framework, we still don’t have sufficient information to determine copyright’s optimal scope perfectly. My own sense is that we have sufficient information to design copyright well enough. Yet, if one accepts the argument that we can’t tell whether the additional works that broader copyright may bring forth are more or less valuable than the alternative uses to which those resources would otherwise be devoted, that lack of empirical evidence inevitably leads to a simple answer as to the optimal scope of copyright protection: none. This is because copyright imposes real costs. If it does not generate equally real welfare gains, it cannot be justified. The only net welfare gain copyright offers is encouraging the production of additional original works when that represents a more valuable use of the available resources than the alternative to which those resources would otherwise be devoted. If we can’t tell which use is more highly valued, then copyright has no welfare gains to offer and should either be abolished or be recognized as a poorly designed charitable mechanism for redistributing wealth from society to copyright owners. See Tom W. Bell, Authors’ Welfare: Copyright as a Statutory Mechanism for Redistributing Rights, 69 BROOK. L. REV. 229, 231-32 (2003).

\textsuperscript{59} As a general rule under trade secret law, a successful plaintiff is entitled to a “lead-time” injunction—an injunction that lasts only for the time it would have taken the defendant to discover the secret lawfully, either by reverse engineering, independent development, or otherwise. See, e.g., K-2 Ski Co. v. Head Ski Co., 506 F.2d 471, 474 (9th Cir. 1974). In their treatise on trade secret law, Roger Milgrim and Eric Bensen cite a number of illustrative cases applying this principle. See ROGER M. MILGRIM & ERIC E. BENSEN, 4 MILGRIM ON TRADE SECRETS § 15.02[1][d] (2014). Some of these cases deny injunctive relief on the grounds that the lead-time period had already expired; some grant permanent injunctions. Of the twenty cases they cite that grant an injunction for a specific time, the duration of the injunctions ranges from 3 months to 10 years, with a mean of 2.5 years, a median of 2 years, and a mode of 3 years. Id. § 15.02[1][d] n.20.

\textsuperscript{60} See Mark A. Lemley, Rational Ignorance at the Patent Office, 95 NW. U. L. REV. 1495, 1504 (2001) (presenting in Table 3 data showing that patentees pay the maintenance fees due eleven years after the patent issues for less than forty percent of the patents issued). Moreover, even while a patent remains in force, it does not usually preclude competitive entry. See Edwin Mansfield, Mark Schwartz & Samuel Wagner, Imitation Costs and Patents: An Empirical Study, 91 ECON. J. 907, 913 (1981) (“Contrary to popular opinion, patent protection does not make entry impossible, or even unlikely. Within 4 years of their introduction, 60% of the patented successful innovations in our sample were imitated.”).

\textsuperscript{61} Alternatively, under this approach, one might argue that patent protection should be longer but, given the relative political economies involved, the patent term is likely to be far closer to socially optimal than the copyright term.

\textsuperscript{62} See Lunney, supra note 24, at 626 (“To the extent the relative copying advantages depicted in Tables 6 and 9 represent the relative copying advantages available to authors generally, the empirical evidence tends to establish that copyright should protect literary works only against exact or near-exact duplication.”).
Adopting more realistic assumptions and recognizing the need to account for the imperfections inevitably present elsewhere in the economy thus suggests a far narrower optimal scope for copyright than does a partial equilibrium approach. Yet, a more careful examination of the second usually unspoken assumption—that consumers will free ride and substitute a less expensive or free, unauthorized copy for an authorized copy whenever they can—may go even further. It may suggest that there is no need for copyright at all.

When it comes to the question of whether a rational, self-interested consumer will either (i) pay for access or (ii) free ride, the consumer has two conflicting desires that she must reconcile. First, she enjoys the work and wants it to exist, so she wants the author to receive sufficient payment to ensure the work’s creation. Second, while wanting to ensure the work’s creation, our consumer also wants to pay as little for access to the work as possible. In sum, our music consumer wants the work for free, but she knows that if everyone free rides, the work will not be created.

We can analyze how a consumer will resolve this dilemma using a game theoretic framework and the principles of Nash equilibrium. When we do, we find a mixture of paid access and free riding that, in the absence of copyright, may lead to suboptimal incentives, but may not. It depends on our starting assumptions. In economic theory and in the economy, there are some public goods that markets can produce efficiently, without the need for specific government intervention. For our purposes, the relevant distinction is between continuous public goods, where the choice is between more or less of the public good at issue, and discrete public goods, where the choice is between having or not having the public good. With respect to music and works of authorship more generally, if we assume that the relevant market is for original works generally and analyze the issue as one involving a continuous public good, then we find the familiar, underproduction result. While consumers will pay something for access to the original works, there will be too little paid access and too much free riding and as a result, the market will produce too few original works in the absence of copyright. On the other hand, if we assume that the relevant market is for a particular work of authorship—a particular song or book or movie, rather than just more songs, books, or movies, and analyze the issue as one involving discrete public

63. As John Nash explained it: “an equilibrium point is . . . such that each player’s . . . strategy maximizes his payoff if the strategies of the others are held fixed.” John Nash, Non-Cooperative Games, 54 ANNALS OF MATHEMATICS 286, 287 (1951); see also John F. Nash, Equilibrium Points in n-Person Games, 36 PROC. NAT’L ACAD. SCI. 48, 48-49 (1950) (introducing his equilibrium concepts).

64. For a formal partial equilibrium proof of this result, see Lunney, supra note 55, app. at 449-51.
goods, then we reach a quite different conclusion. At the robust Nash equilibria, while there may be free riding, consumers will nevertheless pay enough to ensure the work's creation. As a result, if consumers view individual works, rather than works generally, as the relevant market, then the market will produce an optimal supply of original works even in the absence of copyright.

Consider a simple subscription model. There are N consumers who each have some reservation value for a work, \( V_i \). To ensure creation and dissemination of the work, the author must receive some reservation cost, \( C \). Returning to a partial equilibrium approach, for the sake of simplicity, assume that the work is worth more than it costs, or \( \sum V_i \geq C \), and therefore welfare would be improved if the work is created. Each consumer can make a binding commitment to pay some price for access to the work, \( P_i \), or can choose to free ride. The work will be created if the sum of the prices consumers promise to pay exceeds the cost of the work: \( \sum P_i \geq C \). Otherwise, the work will not be created. Consumers pay only if the work is created.

Given this set-up, we find two types of Nash equilibria. In the first, the sum of the promised prices exactly equals the author's reservation cost: \( \sum P_i = C \). While some free riding may occur in these equilibria, enough consumers pay for access to cover the author's reservation cost. As a result, the work is created, and the market reaches an efficient and Pareto optimal outcome. In the second, the sum of the promised prices are insufficient to cover the author’s reservation cost, \( \sum P_i < C \), and no single individual can increase their price sufficiently to make up the difference, \( C - \sum P_i > V_i - P_i \) for \( i=1, N \). For these equilibria, too much free riding occurs, and the promised payments fail to cover the author’s reservation cost. As a result, the work is not created, and the market fails to achieve the efficient outcome.

65. See Lunney, supra note 56, at 15.


67. See Lunney, supra note 56, at 10-16. Instances where the sum of the prices exceeds the author’s reservation costs are not Nash equilibria, because each consumer would want to reduce her promised price until the sum of the prices exactly equals the author’s reservation costs.

68. Id. at 12-13.

69. If a single individual could make up the difference, then the individual would be better off doing so, and we would move to the first type of Nash equilibria.
While these are both Nash equilibria, only the efficient equilibria, where the work is created, are robust. At one of the efficient Nash equilibria, for each and every consumer, any change in strategy would make the consumer worse off. Of course, even at these equilibria, each consumer would still prefer to pay less if they could, but they cannot and so they will not. Attempting to pay less (or to pay nothing) would mean that the work would not be created and at these Nash equilibria, each consumer prefers to have the work created, given the price that they are paying, than to pay nothing and not have the work at all.

In contrast, with respect to the inefficient Nash equilibria, where the work is not created, none of them are robust. At any one of the inefficient equilibria, a consumer has nothing to gain by changing her strategy, but she has nothing to lose either. At these Nash equilibria, the work will not be created and nothing will be paid. As a result, consumers who did not promise their full reservation value are indifferent between their current promised price and a slightly higher or lower promised price. Moreover, such a consumer will actively prefer any of the Nash equilibria where the work is created to any of the Nash equilibria where it is not. Given that preference, if a consumer has any uncertainty as to what other consumers will bid, a consumer who has promised a price less than his or her reservation value has an incentive to increase his or her promised price. By bidding somewhat more, a consumer can increase the likelihood that the work will be created. Taken together, this means that the inefficient Nash equilibria are not robust. No consumer has any incentive to stay at any of them and if they could find a way to move to one of the efficient Nash equilibria, they would.

Given this and contrary to longstanding belief, the market may well be capable, even in the absence of copyright, of producing original works of authorship efficiently. So long as consumers desire specific, individual works of authorship and therefore treat original works as discrete public goods, the massive government intervention that copyright represents may prove entirely unnecessary.

As is so common with economic modeling, our starting assumption thus dictates our result. An assumption that the relevant market is for works generally leads us to the continuous public goods model and the conclusion that the market will invariably produce too few original works. In contrast, an assumption that consumers are looking for a specific work leads us to the discrete public goods model and

---

70. For a discussion of the concepts economists use to sort Nash equilibria, see ERIC VAN DAMME, REFINEMENTS OF THE NASH EQUILIBRIUM CONCEPT (1983).

71. I am employing the conventional tie breaker, such that having the work is worth just slightly more to the consumer, even at her reservation value, than not having the work.
the opposite conclusion: For such discrete public goods, the market can achieve an efficient, Pareto optimal outcome and produce an optimal supply of original works. The question then becomes whether the continuous or discrete public goods analysis better matches the real world characteristics of original works. This in turn depends upon whether consumers: (i) consider any given original work sufficiently unique to constitute its own market; or (ii) consider each original work merely an indistinguishable widget that together with other original works (also indistinguishable widgets) form a broader market for works generally. Which model better describes the actual market or markets for original works will thus depend entirely on consumer preferences and may vary for different types of works.

Yet, to the extent consumers consider an original work sufficiently unique to constitute its own market, we may not need copyright at all. Instead, all we may need are market mechanisms that (i) enable consumers to commit to purchasing access to a work; and (ii) help consumers reach one of the efficient Nash equilibria. While such mechanisms may have been difficult to implement in an analog world, we have already seen several practical implementations of such mechanisms on the Internet, including Kickstarter, Stephen King's marketing of the novella, *The Plant*, and Nine Inch Nails's release of tracks from its *Ghost I-IV* album.

Economic theory thus reaches somewhat inconsistent conclusions on the underlying need for copyright and on how to define its proper limits, depending on the assumptions with which we begin. We need not rely on theory alone, however. The rise of file sharing provides a rare opportunity to test copyright’s fundamental premises and the competing predictions of economic theory against economic reality. In the next section, we explore how the rise of file sharing has affected creative output in the music industry.

---

72. As I have discussed elsewhere, lighthouses represent a discrete public good and were initially funded through a subscription model. See Lunney, *supra* note 56, at 22.

73. Kickstarter is a website that allows individuals to propose projects and solicit donations to support the projects. It is a subscription-based model where the donations are made, and the project only proceeds if a specified level of support is reached. Launched on April 21, 2009, Kickstarter has enabled individuals to raise an estimated $1 billion for their projects. See Seven Things to Know About Kickstarter, KICKSTARTER, https://www.kickstarter.com/hello (last visited Feb. 11, 2015). Musicians have used Kickstarter to pay for the production costs for a new album and for tours. For example, in 2012, Amanda Palmer raised more than one million dollars from more than 17,000 individual donations for a tour and a new album. See Alison Fensterstock, *Making Her Own Way: Do-It-Yourself Singer Amanda Palmer Uses 1990s Ingenuity and 21st Century Savvy to Finance Her Rise to Fame*, THE TIMES-PICAYUNE, Sept. 15, 2012, at C-1.

B. Testing Theory Empirically: File Sharing and Music Output

We are all familiar with file sharing’s basic story. Since Napster opened its virtual doors in 1999, widespread consumer copying and distribution of copyrighted works through file sharing services has become the new reality. Copyright owners have tried to stop it, of course. They have sued both file sharing services, such as Napster, Aimster, and Grokster, and individual file sharers. While winning the vast majority of these battles, copyright owners have just as surely been losing the war. Despite their lawsuits and their educational campaigns, file sharing has become remarkably widespread. For virtually every copyrighted song, television program, and movie that exists, a consumer, without much effort, can obtain her own unauthorized copy for free by file sharing. As a formal matter, copyright continues to provide an extremely high level of protection; as a practical matter, the effective level of protection copyright provides to original works has fallen radically. Given this sharp reduction in copyright protection, the natural question becomes: How has it affected creative output?

To begin our examination of this issue, Figure 2 presents an estimate of the file sharing traffic on the Internet in North America, from Cisco’s Visual Networking Index for 2008 and 2014.

---


76. See Sony BMG Music Entm’t v. Tenenbaum, 660 F.3d 487, 489-90 (1st Cir. 2011) (affirming finding of copyright infringement against an individual for file sharing); Capitol Records, Inc. v. Thomas-Rasset, 799 F. Supp. 2d 999, 1001-02, 1016 (D. Minn. 2011) (affirming finding of copyright infringement for uploading twenty-four songs through a file-sharing program, but reducing damages award to $54,000).

As Figure 2 reflects, file sharing traffic in North America in 2012 amounted to roughly eight hundred petabytes per month. Just to give a sense of scale to this issue, the typical music CD contains 800 megabytes of data. One step up from a megabyte is a gigabyte. A gigabyte is one thousand megabytes, and a typical DVD contains 4 gigabytes of data. A petabyte is one million gigabytes. At eight hundred petabytes per month, the current rate of file sharing traffic represents approximately 200 million DVDs or 1 billion CDs, copied each month. Compare that to the roughly 139 million albums that I estimated U.S. consumers made through file sharing on Napster in September 2000, and we can see that, despite the copyright industries’ “victories” over file sharing, file sharing has increased substantially over the past twelve years. While not all of this traffic represents the unauthorized copying and distribution of copyrighted works, estimates suggest that the vast majority of it does.

Interestingly, in its two most recent indices, released in May 2013 and June 2014, Cisco projects that file sharing traffic in North America will grow much more slowly from 2012 to 2018 than it did from

---


79. Cf. Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 545 U.S. 913, 922 (2005) (noting that a study by an expert for the plaintiff showed that “nearly 90% of the files available for download on the FastTrack system were copyrighted works”).

Instead of continuing to grow by fifteen to twenty percent annually, as Cisco had consistently projected in its 2008, 2009, 2010, 2011, and 2012 indices, in its 2013 and 2014 indices, Cisco projected that file sharing traffic will grow by only seven to nine percent annually in North America from 2012-2018. It is not clear what is behind this slowdown. In its 2013 and 2014 indices, Cisco projects a slowdown in the growth rate for file sharing traffic over the next four years—but the slowdown is not limited to just North America. The slowdown is projected worldwide. Indeed, for two of its regions, Africa and the Middle East, and Western Europe, Cisco projects that file sharing traffic will actually start declining, at respective rates of twelve and two percent annually, over the next four years. Given that the slowdown is worldwide, the slowing growth rate does not appear to be the result of a legal intervention by any particular country. It may simply be that file sharing is running its inherently self-limiting course and reaching an equilibrium. In a finite world, nothing can continue to grow geometrically indefinitely. Yet, regardless of whatever is causing the slowdown in the growth of file sharing traffic, the slowdown itself tends to diminish the need for further legal intervention to address the file sharing issue.

In any event, unable to stop file sharing, the effective level of copyright protection provided to original works has fallen dramatically since 1999. As file sharing has grown and the de facto level of copyright protection provided has fallen, the music industry, in particular, has been hard hit. With the rise of file sharing, the music industry has seen revenue from record sales decline steadily and sharply. To illustrate, Figure 3 presents the RIAA’s total dollar value for music shipments, in all formats, whether physical or electronic, from 1973 to 2013. In order to account for inflation, shipments are in constant 2013 dollars.


82. This data is from the RIAA Shipment Data. Shipment Database, supra note 4 (after subscribing as a member, select shipment data from year 1973 through 2013 to download data). The year 1973 is as far back as the RIAA data goes.
As Figure 3 reflects, since Napster opened its doors, the RIAA reports that shipments of music have fallen from a peak of $20.4 billion in 1999 to only $7 billion in 2013. This is a fall of some $13 billion or 65.7 percent. Such a fall is not entirely unprecedented. A similar peak and fall occurred from 1978 through 1982, when shipments peaked at $14.8 billion (in constant 2013 dollars) before falling to $8.8 billion in 1982—a fall of some $6 billion or 40.7 percent. Presumably, not even the music industry would contend that file sharing caused this initial fall in music sales. Rather, this initial fall was likely due to difficulties in the economy generally from 1980 through 1982. Because music is a luxury good, spending on it can fall quite rapidly when unemployment rates rise or per capita income falls.

Yet, even if not entirely unprecedented, the decline in music shipments that follows Napster and tracks the rise of file sharing is both sharper and lasts longer than the decline during the early 1980s. Undoubtedly, difficulties in the economy generally, particularly after the start of the Great Recession in 2008, contribute to the post-Napster decline in music shipments. But file sharing may also have played a role. Existing economic studies disagree as to whether and if so, to what extent, file sharing may have contributed to this decline.83 Yet, I am perfectly prepared to accept that file sharing is responsible, directly or indirectly, for some part of this decline.

83. See sources cited supra note 6.
The relevant question, however, is not whether file sharing caused this decline in record sales, but how this decline in record sales affected the creation of new music. The Constitution gives Congress the authority to enact copyright law “to promote the Progress of Science...” This is the sole standard against which copyright law must measure itself. In defining this standard, the Court has explained that promoting the “Progress of Science” encompasses two legitimate ends: (i) encouraging the creation of new works; and (ii) encouraging the dissemination of existing works. File sharing undoubtedly promotes the widespread dissemination of existing works. The concern has been that it will discourage the creation of new works because copyright owners are not directly compensated for the unauthorized copies of their works distributed through file sharing networks. Given that file sharing plainly serves to promote broader dissemination, the question becomes whether there is equally clear evidence that file sharing discourages the creation of new works. Before examining this question directly, I should acknowledge that record sales are not the only source of revenue that supports the creation of new music. Sync licenses, endorsement deals, concert ticket sales and associated merchandising, and public performance royalties, just to name a few, provide revenue streams that support the creation of new music, as well. Many of these are complements to the sale of recorded music and so may serve as a means for recapturing all or part of the revenue lost as a result of declining music sales.

While possible in theory, in reality, growth in the revenues associated with complementary products, such as live performances, has not done much to mitigate the decline in music sales. While we do not have data on all of the revenue sources available to artists and songwriters, two of the main alternative revenue sources, concert revenue and public performances royalties, are available, and they have not grown sufficiently to offset the revenues lost from falling record sales. As reported by Pollstar, revenue from concerts in North America grew from $2.07 billion in 1999 (adjusted to 2012 dollars) to $4.3 billion in 2010.  

---

86. A synchronization license, or “sync” license, is required when a movie, television program, or advertisement wants to include a musical work and synchronize the music to the visual images. See Leadsinger, Inc. v. BMG Music Publ’g, 512 F.3d 522, 527 (9th Cir. 2008).
87. See Roy Trakin, The Concert Puzzle: Can the Live Concert Industry Rebound from a Difficult 2010?, GRAMMY.COM (Mar. 25, 2011, 6:00 AM), http://www.grammy.com/news/the-concert-puzzle (“Over the last decade, as the recorded music industry contracted at a double-digit annual rate, the live concert touring business continued to grow, with worldwide ticket revenues rising from $1.5 billion in 1999 to $4.6 billion in 2009, according to Pollstar.”). Adjustments to 2012 dollars were performed using the Bureau of Labor and
billion in 2012.\footnote{Pollstar’s Top 200 North American Tours, POLLSTAR (Jan. 7, 2013, 5:01 PM), http://www.pollstar.com/news_article.aspx?ID=803803 (“With all 2012 box office reports for North America totaling more than $4.3 billion, the Top 100 tours accounted for more than half that amount, generating $2.52 billion.”).} Similarly, royalties paid by ASCAP and BMI for the public performance of musical works grew from $1.29 billion in 1999 (again, adjusted to 2012 dollars) to $1.58 billion in 2012.\footnote{ASCAP Reports 2012 Financials, ASCAP (Mar. 4, 2013), http://www.ascap.com/press/2013/0304-ascap-reports-2012-financials.aspx (“The American Society of Composers, Authors and Publishers (ASCAP), the worldwide leader in performing rights and advocacy for music creators, today announced that it distributed over $827 million in royalties to its songwriter, composer and publisher members in the calendar year ended 2012, a slight increase over 2011.”); Ed Christman, BMI Releases Results, Revenue Down 3.5% Due to Radio Rate Settlement, BILLBOARD (Sept. 20, 2012, 6:02 PM), http://www.billboard.com/biz/articles/news/publishing/1083703/bmi-releases-results-revenue-down-35-due-to-radio-rate (“During the recently completed year, BMI distributed $749.8 million in royalties, down 5.8% from the $796 million paid out in the prior year.”); Tamara Conniff, ASCAP Composes Record Year, HOLLYWOOD REP., Feb. 14, 2001 (“The American Society of Composers, Authors and Publishers’ revenue reached a record high of $576 million last year, up $41 million from 1999, the company said Tuesday. Domestic and foreign royalty distribution to its members totaled $479.1 million. Domestic distribution increased 12% from 1999, ASCAP chief John LoFrumento said.”); Sherman Friedman, BMI Launches Online Song Registration, NEWSBYTES, June 19, 2000 (“BMI currently distributes about $500 million in worldwide royalties from all media to its members on an annualized basis.”).} Since 2003, SoundExchange has also begun paying royalties for the public performance of sound recordings and has paid out approximately $462 million in 2012.\footnote{SoundExchange, SOUNDEXCHANGE ANNUAL REPORT FOR 2012, at 7 (2013), available at http://www.soundexchange.com/wp-content/uploads/2013/06/2012-Annual-Report-06-13-13.pdf (“In 2012, SoundExchange’s gross distributions were approximately $462 million.”).} While these revenue sources grew faster than the rate of inflation, the total increase in revenue available from these alternative sources from 1999 to 2012, is not enough to offset the $13 billion decline in revenue from record sales over that same period.\footnote{Cf. Felix Oberholzer-Gee & Koleman Strumpf, File Sharing and Copyright, in 10 INNOVATION POLICY AND THE ECONOMY 21, 23 (2010), available at http://www.nber.org/chapters/c11764.pdf (suggesting that sale of complementary goods has largely offset loss of record sale revenue, but including iPod sales and failing to account for inflation by using a nominal, rather than constant, dollar analysis).}

This brings us to the heart of the matter: how has this decline in revenue affected creative output in the music industry? For four hundred years, copyright’s fundamental premise has been that more revenue will lead to more and better creative works. If this premise is accurate, then the sharp decline in music industry revenue since 1999 should have led to fewer or lower quality songs. Surprisingly, perhaps, given copyright’s premise, it has not.


88. See ASCAP Reports 2012 Financials, ASCAP (Mar. 4, 2013), http://www.ascap.com/press/2013/0304-ascap-reports-2012-financials.aspx (“The American Society of Composers, Authors and Publishers (ASCAP), the worldwide leader in performing rights and advocacy for music creators, today announced that it distributed over $827 million in royalties to its songwriter, composer and publisher members in the calendar year ended 2012, a slight increase over 2011.”); Ed Christman, BMI Releases Results, Revenue Down 3.5% Due to Radio Rate Settlement, BILLBOARD (Sept. 20, 2012, 6:02 PM), http://www.billboard.com/biz/articles/news/publishing/1083703/bmi-releases-results-revenue-down-35-due-to-radio-rate (“During the recently completed year, BMI distributed $749.8 million in royalties, down 5.8% from the $796 million paid out in the prior year.”); Tamara Conniff, ASCAP Composes Record Year, HOLLYWOOD REP., Feb. 14, 2001 (“The American Society of Composers, Authors and Publishers’ revenue reached a record high of $576 million last year, up $41 million from 1999, the company said Tuesday. Domestic and foreign royalty distribution to its members totaled $479.1 million. Domestic distribution increased 12% from 1999, ASCAP chief John LoFrumento said.”); Sherman Friedman, BMI Launches Online Song Registration, NEWSBYTES, June 19, 2000 (“BMI currently distributes about $500 million in worldwide royalties from all media to its members on an annualized basis.”).


As Figure 4 illustrates, the quantity of new albums released in the United States is up substantially since Napster’s debut. Using a consistent methodology, SoundScan estimates that the number of new albums released increased from just under forty thousand albums in 1999 to a peak of over one hundred thousand albums in 2008. With the onset of the Great Recession in 2008, the number of new albums fell back in 2009-2012, but even so, the number of new albums released in 2012, at just under eighty thousand, nearly doubles the pre-Napster output.

While the increasing number of albums released suggests that the revenue decline has not affected music output in terms of quantity, it does not account for a possible decline in quality, nor does it foreclose the possibility that even more albums might have been released but-for the decline in revenue. Of course, to a considerable extent, such arguments are either entirely disingenuous or the product of wishful thinking. When the Great Recession hit in 2008 and new home and automobile sales fell precipitously, there was a corresponding, and immediate fall in new housing starts and automobile production in the United States.92 For either industry, we did not need to argue

---

92. According to the U.S. Census Bureau, new housing starts fell from over twenty million units a year in 2004, 2005, and 2006, to only 10.2 million units in 2008, and then to only 6.6 million units in 2009—a decline of more than 50 percent. See New Residential Construction: Seasonally Adjusted Annual Rate for Housing Units Started, U.S. CENSUS BUREAU, http://www.census.gov/econ/currentdata/ (last updated Mar. 26, 2014) (select “New Residential Construction”; then press “submit”; then under “Select Industry or Category” select “Annual Rate for Housing Units Started”; then press “Get Data”). Similarly, the International Organization of Motor Vehicle Manufacturers (OICA) reports that automobile production in the United States fell from over 10 million units in 2004, 2005, and 2006 to only 5.7 million units in 2009. Production Statistics, OICA, http://www.oica.net/
that while housing starts or automobile production remained constant, the size, quality, and features of the homes or cars had fallen. Nor did we need to argue that output in either sector remained high, but output would have been even higher but for the recession. The sort of sharp and immediately measurable decline in output that these industries experienced reflects how a competitive and well-functioning market responds to a sharp decline in revenue. In a competitive market, firms are already operating at the margins of profitability; they cannot simply absorb a sharp decline in revenue and maintain production. That the music industry was able to do so suggests that the music industry may not operate in a comparably competitive and well-functioning market.

In a just completed empirical study, I have nonetheless attempted to examine these issues directly and to account for the possibility that: (i) music quality might have fallen as a result of the revenue decline; or (ii) music output might have been higher but for the revenue decline. In the study, I treat the rise of file sharing and the parallel fall in music industry revenue as a natural experiment in radically reduced copyright protection. To explore the relationship between copyright protection, revenue, and high quality creative output, I created a hand-coded data set for the songs that appeared in the top fifty of the Billboard Hot 100 in the first week of each month for each year from 1985 through 2013. I focused on songs that appeared in the top fifty of the Hot 100 in order to control for quality.

Both before and after file sharing, reaching the top fifty of the Hot 100 provided some evidence of a song’s ability to satisfy the musical preferences of music consumers. At a minimum, a new song will hit the top fifty of the Hot 100 only if consumers prefer it to the preexisting songs otherwise available. Appearance in the top fifty thus provides a reasonable proxy for quality that should be consistent across the pre- and post-file sharing eras.

In the data set, for each of the 17,400 songs in the study, I coded, inter alia: (i) whether the song was performed by a New Artist, in that it was the artist’s first appearance in the Top 50 and occurred

---

94. See Lunney, supra note 7.
96. While listeners may have a preference for “new,” as well as “good,” music, there is no evidence that this preference has changed from the pre- to the post-file sharing era.
with, or before, the artist’s first single on the artist’s first studio album; and (ii) whether the song was a New Song, in that it had not previously appeared in the Top 50 and was not a cover. Once I had identified the New Artists for each year, I went back and created a second data set. In the second data set, for each New Artist, I determined: (i) the number of Top 50 chart appearances each New Artist had during the first ten calendar years of their career; and (ii) the number of different songs with which each New Artist hit the Top 50, again during the first ten calendar years of their career. Because this second data set covered an artist’s first ten years, it covers a more limited time period and covers only those New Artists who first appeared in the Top 50 from 1985-2005.97

After collecting the data, I used regression analysis to isolate and define any statistically significant correlations between revenue, file sharing, and music output. By using regression analysis, I was able to isolate the relationship between revenue and music output, regardless of other changes that may have been occurring in the economy generally or in the music industry specifically, such as the declining cost of album production. The regression analysis established that the sharp decline in revenue from music shipments following the rise of file sharing was associated, if we hold all else constant: (i) with fewer new artists entering the market; but (ii) also with more hit songs, on average, by those new artists who did enter. The first marginal effect, that the number of new artists declined as revenue fell, is entirely consistent with copyright’s traditional incentives story. As revenue fell, an artist’s expected earnings from pursuing a career in music also fell, and so, at the margins, some individuals decided to devote their talents elsewhere.

The second marginal effect, that existing artists produced more music as revenue fell, is not, however, consistent with copyright’s traditional story.98 That existing artists produced more music as revenue fell suggests that, during the 1990s, the returns that copyright provided musical were too high. In the study, I suggest that given copyright’s effective scope at that time, our most popular artists were receiving an incentive for their music far in excess of their reservation price. These excess incentives had gotten so high that they were pushing our most popular artists onto the backward bending portion of the labor supply curve. Finding themselves on the backward-

97. For the New Artists from 2005, it covers only the first nine calendar years of their careers.

98. One could suggest that given the decline in new artists, it was inevitable that preexisting artists would fill the available spaces, but that fundamentally misunderstands the Billboard Hot 100. The chart does not guarantee space to a certain number of new songs each year. Rather, a new song can earn a place on the Hot 100 only if it is more popular than, and hence can displace, existing songs on the chart.
bending portion of the labor supply curve, our most popular artists chose to substitute leisure for work, and so produced less music.

As file sharing became an option, it sharply reduced these excess incentives. With the rise of file sharing, revenues for creating and disseminating any given song fell.99 As a result, fewer artists found themselves on the backward portion of the labor supply curve, and so each new artist produced, on average, more hit songs, than each new artist did when revenues were high. In short, when revenues from music sales fell with the rise of file sharing, those individuals who decided to devote their talents to music began to produce more of it.

We can see this not only in the regression results, but in the data itself. From 1985 through 1999, a fifteen-year period, the top new artists were Sean Combs, who appeared on twenty-four different Top 50 songs in the first ten calendar years of his career; Mariah Carey, who had twenty such appearances; and Whitney Houston, who also had twenty such appearances. In contrast, with the rise of file sharing and the fall in revenue from record sales, from 2000 through 2006, the top new artists were Rihanna, who had thirty-five such appearances in the first nine years of her career; Taylor Swift, who had twenty-nine such appearances in the first eight years of her career; Ludacris, who had twenty-nine such appearances in the first ten years of his career; Akon, who had twenty-six such appearances; and 50 Cent, who had twenty-four such appearances. In other words, in just seven years during the low revenue era of file sharing, we had five new artists who each had as many or more Top 50 appearances than the top new artist from the preceding fifteen, high revenue years. The regression results establish that this is not a coincidence or random chance. There was a statistically significant correlation between higher revenue from music shipments and fewer hits by each new artist.

Although counterintuitive, this result is entirely consistent with the backward-bending labor supply curve that economic theory pre-

dicts and that empirical studies have verified.\textsuperscript{100} When revenues were high, new artists did not have to work as hard to achieve the lifestyle they desired, and so they chose at the margins to substitute leisure for work. As revenues fell, new artists had to work harder to achieve the lifestyle they desired and so, at the margins, they choose to substitute work for leisure. As a result, as revenues fell, output from those new artists increased.

Moreover, the case study not only isolated and identified these two marginal effects, it used regression analysis to calculate their relative magnitude precisely. As it turns out, for the period studied, the second marginal effect was larger than the first. Any given decline in revenue from music shipments had a larger marginal effect on output from existing artists than it did on the supply of new artists. As a result, the decline in the dollar value of music shipments since file sharing began was associated with a net increase in the number of new hit songs, holding all else constant.\textsuperscript{101} We got fewer new artists, but more hit songs from the new artists we had. For the music industry, the rise of file sharing and the associated decline in revenue thus meant the creation of more new hit songs.\textsuperscript{102} Indeed, file sharing increased music output precisely because it decreased music industry revenue.

Even though file sharing thus led to both more new music and far broader dissemination of existing music, copyright owners have, since file sharing began, been begging Congress and the courts to wave their magic pens and put a stop it. For the first time, this case study allows us to see precisely how music output would have changed had their efforts proven successful. If file sharing had been

\begin{itemize}
\item \textsuperscript{100} “[O]nly lottery winners experience sudden wealth in a way similar to that of suddenly popular authors today. Studies of lottery winners demonstrate that such large awards sharply reduce, on average, time worked.” Lunney, supra note 9, at 891. As Professor Gregory Mankiw has summarized:

The results from studies of lottery winners are striking. Of those winners who win more than $50,000, almost 25 percent quit working within a year, and another 9 percent reduce the number of hours they work. Of those winners who win more than $1 million, almost 40 percent stop working. The income effect on labor supply of winning such a large prize is substantial.

N. GREGORY MANKIW, PRINCIPLES OF MICROECONOMICS 483 (2d ed. 2000).

\item \textsuperscript{101} See Lunney, supra note 7, at 25 (“The New Song regression estimates that a one million dollar decline in record sales was associated with a 0.00466 increase in the number of new songs appearing in the study’s sample of the top fifty on an annual basis, holding all else constant. If we multiply this coefficient by a $13 billion decline in record sales, the regression results would estimate that this decline would be associated with 60.6 additional new songs entering the study’s sample of the top fifty on an annual basis.”).

\item \textsuperscript{102} For similar results using other measures of music quality, see Joel Waldfogel, And the Bands Played On: Digital Disintermediation and the Quality of New Recorded Music (Carlson Sch. of Mgmt. and Dept of Econ., Preliminary Draft, 2012), available at http://www.ssrn.com/abstract=2117372, and also see Oberholzer-Gee & Strumpf, supra note 91, at 20 (noting that book publishing, movie production, and music output have all risen substantially since file sharing began).
\end{itemize}
stopped and revenue from music shipments had remained at their peak 1999 levels, we would have had somewhat more hit songs because higher revenue would have encouraged more new artists to enter the field. But, we would also have had fewer hit songs because higher revenue would have encouraged those new artists who did enter the field to devote more time to leisure. Because the second marginal effect would have outweighed the first, we would, in the end, have had significantly fewer new songs in the Top 50 of the Billboard Hot 100 had file sharing been stopped and had revenues from music shipments remained high. Using the basic regression results presented in Table 2 of the study, Figure 5 presents a trend chart showing precisely how much less music we would have had if file sharing had been stopped.

Figure 5. New Music: Actual and Projected (If File Sharing Had Been Stopped and Revenue Had Remained High)

As Figure 5 illustrates, had the recording industry successfully stopped filed sharing and kept their revenue at peak 1999 levels, we would have seen significantly fewer hit songs since 2001. While radically counterintuitive, these results unmistakably establish something many of us in academics have long suspected. Copyright has gotten radically overbroad. It lasts far too long and provides far broader protection than is necessary or desirable to achieve its constitutional purposes. In the 1990s, copyright had gotten so broad that

103. Lunney, supra note 7, at 23.

104. To calculate the projected number of New Songs, I multiplied the coefficient on music shipments from the regression analysis by the difference in music sales had sales remained constant at the peak 1999 level rather than fallen for each year from 2001 through 2013, and I subtracted the resulting change in new song production from the actual number of new songs that appeared in the sample for each year.
it was generating incentives for our most popular artists so far in excess of their reservation price that we were, perversely, getting less music from them. Had the record industry’s efforts to stop file sharing proven successful, we would have continued vastly overpaying our most popular artists for their music. And had we continued to overpay them, our most popular artists would have shown their appreciation by giving us fewer hit songs in return. The net result: Had revenues remained high, fewer new hit songs would have been produced over the last ten years.

File sharing provided a necessary and desirable corrective. By allowing consumers to copy and distribute original music freely, outside the restrictive strictures of copyright, the rise of file sharing helped reduce the extent to which copyright was overprotecting original works. By reducing the effective level of protection copyright provided, file sharing reduced the excess incentives for creating new music copyright would otherwise have provided. By reducing these excess incentives, file sharing led to the production of more hit songs.

Copyright’s fundamental premise that more revenue equals more works is thus not always true. While some copyright may generate more original works, too much copyright and too much revenue can lead to fewer works, as the regression results establish was happening in the music industry before the rise of file sharing. Under copyright’s traditional and constitutionally sanctioned justification, there thus appears no plausible basis for broadening copyright in an attempt to control file sharing, at least, for music. Given how broad copyright is today, broadening copyright even further will not lead to more and better works, but instead will lead to fewer and worse. That helps to explain why proponents of broader copyright have rushed to embrace other arguments, such as increased employment, to support their demands for broader copyright. In the next section, we turn to this alternative argument and attempt to answer the question of whether increasing revenues to and jobs in the copyright industries can provide an alternative justification for broader copyright. As we do, we must take care to evaluate the argument on its own merits and not allow copyright’s traditional justification to color our thinking. We will therefore assume, consistent with the results from my case study of file sharing and music output, that any increase in revenue to the copyright industries will not lead to increased creative output.\textsuperscript{105}

\textsuperscript{105} Even if you are reluctant to embrace the results of my case study, it remains worthwhile to determine whether the jobs argument can support broader copyright independently of copyright’s traditional more or better works storyline.
IV. IN THE ABSENCE OF MORE OR BETTER WORKS, CAN MORE JOBS ALONE JUSTIFY BROADER COPYRIGHT?

If broader copyright does not lead to more or better works of authorship, and it may in fact lead to fewer, as the evidence suggests, then we cannot justify broader copyright by the additional jobs it may create. While broader copyright can increase copyright industry revenue, and may thereby lead to more jobs in the copyright industry, it does so by redistributing wealth from consumers to the copyright industry. Such a redistribution generates no net welfare gain on its own. It simply takes money from one group and gives the same money to another. Indeed, given the transaction and administrative costs such a redistribution would entail, any such redistribution necessarily generates a net welfare loss. Moreover, the possibility that such additional revenue may generate additional employment in the copyright sector also cannot justify such a redistribution. As Frederic Bastiat explained more than a century ago, the sort of redistribution that broader copyright achieves generates no net stimulus for the economy. Any additional jobs broader copyright might create in the copyright sector would simply displace employment elsewhere in the economy.

A. Of Broader Copyright and Broken Windows

As discussed in the introduction, the basic thrust of the mercantilist argument for SOPA and PIPA—that these measures will increase revenue to and hence jobs in the copyright industries—represents an example of Bastiat’s Broken Window Fallacy. Just as a broken window may yield additional revenue to glaziers by taking it from other sectors of the economy, broader copyright may similarly yield additional revenue to copyright owners, but it does so only by taking it from other sectors of the economy. While we can readily see and account for the additional jobs we hope for in the copyright sector, we must also account for the loss in jobs—not so readily seen, but just as real—elsewhere. When we do, we find that broader copyright generates no net stimulus for the economy as a whole.

Of course, the broken window analogy is not perfect. Copyright is trying to create something of value—new or better works of authorship—not break a window. Nonetheless, the need to evaluate the trade-off between different uses of scarce resources remains. In his essay, Bastiat addressed this point directly in discussing state funding for the arts. Just as supporters of SOPA and PIPA insisted that jobs justified congressional action, so too in Bastiat’s day, supporters

106. BASTIAT, supra note 13, at 72.
107. Id. at 87.
of a proposed government subsidy for theatres in the amount of sixty thousand francs also touted jobs as the justification:

“The economical question, as regards theatres, is comprised in one word – labour. . . . The theatres in France, you know, feed and salary no less than 80,000 workmen of different kinds; painters, masons, decorators, costumers, architects, &c., which constitute the very life and movement of several parts of this capital, and on this account they ought to have your sympathies.”

Or

“The pleasures of Paris are the labour and the consumption of the provinces, and the luxuries of the rich are the wages and bread of 200,000 workmen of every description, who live by the manifold industry of the theatres, and who receive from these noble pleasures, which render France illustrious, the sustenance of their lives and the necessaries of their families and children. It is to them that you will give 60,000 francs.”

Even if we assume, as Bastiat did, that the entire subsidy went to the workers, rather than corrupt government officials, and further assume, again as Bastiat did, that supporting theatre workers is, at some general level, desirable, we still must account for that which is not seen. A subsidy after all does not create wealth that was not there before. “Certainly,” Bastiat wrote, “nobody will think of maintaining that the legislative vote has caused this sum to be hatched in a ballot-box; that it is a pure addition made to the national wealth; that but for this miraculous vote these 60,000 francs would have been for ever invisible and impalpable.” It merely redistributes existing wealth:

[It is clear that the taxpayer, who has contributed one franc, will no longer have this franc at his own disposal. It is clear that he will be deprived of some gratification to the amount of one franc; and that the workman, whoever he may be, who would have received it from him for some service, will be deprived of a benefit to that amount. Let us not, therefore, be led by a childish illusion into believing that the vote of the 60,000 francs may add anything whatever to the well-being of the country, and to national labour. It displaces enjoyments, it transposes wages – that is all.

108. Id. at 92 (quoting M. Lamartine).
109. Id. at 92-93 (quoting M. Lamartine).
110. Id. at 93 (“Yes, it is to the workmen of the theatres that a part, at least, of these 60,000 francs will go; a few bribes, perhaps, may be abstracted on the way. Perhaps, if we were to look a little more closely into the matter, we might find that the cake had gone another way, and that these workmen were fortunate who had come in for a few crumbs. But I will allow, for the sake of argument, that the entire sum does go to the painters, decorators, &e.”).
111. Id. at 93-94.
112. Id. at 94.
While Bastiat acknowledged that one might justify this redistribution by showing that work in a theatre has some value or intrinsic merit other labor lacks, the supporters of theatre subsidies in Bastiat’s day, just as the supporters of SOPA and PIPA today, did not make that argument. They did not suggest that one form of labor was intrinsically more valuable than another. Rather, they pretended that they could, through government action, create additional jobs in one sector at no cost to the rest of the economy. This was as wrong then as it is today. In both cases, the argument “mistak[es] a transfer for a gain.”

Applied to broader copyright, Bastiat’s lesson is simple: Broader copyright does not create jobs; it simply substitutes jobs in the copyright sector for jobs elsewhere. For copyright owners to receive more for their works, consumers must pay more. If consumers pay more for original works, they will necessarily have less for everything else. An extra dollar for copyright owners means a dollar less for everyone else. Broadening copyright so that copyright owners earn millions more means millions less everywhere else in the economy. Those millions may generate more jobs in the copyright sector, but they do so not by employing resources that would otherwise go unused or wasted. Rather, they do so by taking those resources from elsewhere in the economy.

If we pursue the question of whether additional jobs in the copyright sector might be more valuable than jobs elsewhere, we find ourselves back at the central question neoclassical welfare economics asks: Which is the more valuable use of the available resources? If resources are going elsewhere that would be more valuably used to create additional original works, then broadening copyright can make sense. If they are not, then broadening copyright does not make

113. Id. (“Will it be said that for one kind of gratification, and one kind of labour, it substitutes more urgent, more moral, more reasonable gratifications and labour?”).
114. Id. (“There is nothing to prove that this latter class [of theatre workers] calls for more sympathy than the former [class of non-theatre workers]. M. Lamartine does not say that it is so. He himself says, that the labour of the theatres is as fertile, as productive as any other (not more so) . . . .”).
115. While it is true that Bastiat, in his essay, focused on direct government subsidies for the arts, rather than the indirect subsidies that broader copyright provides, the economic consequences of the two are the same. Tom W. Bell, Authors’ Welfare: Copyright as a Statutory Mechanism for Redistributing Rights, 69 BROOK. L. REV. 229, 240-42 (2003); see also Mark A. Lemley, Property, Intellectual Property, and Free Riding, 83 TEX. L. REV. 1031, 1032 (2005) (“If we must fall back on a physical-world analogy for intellectual property protection—and I see no reason why we should—treating intellectual property as a form of government subsidy is more likely to get people to understand the tradeoffs involved than treating it as real property.”). Whether government raises taxes on other employment in order to subsidize the arts directly, or enacts broader copyright so that the prices for original works are higher, “you diminish wages of labourers, drainers, carpenters, blacksmiths, and increase in proportion those of the singers.” BASTIAT, supra note 13, at 94.
116. BASTIAT, supra note 13, at 95.
sense. That broader copyright may generate additional revenue and more jobs in the copyright sector adds nothing to the discussion and is irrelevant. Indeed, it should be obvious that employing more individuals in the copyright sector of the economy in order to produce the same level and quality of creative output is wasteful. If we are going to achieve the same creative output in any event, we might as well use fewer employees to reach that output and employ the additional individuals at issue more productively elsewhere in the economy. In the next section, we explore in more detail the relationship between broader copyright, increased copyright owner revenue, and welfare.

B. Optimal Copyright: Maximizing Welfare Versus Maximizing Jobs

If we ignore Bastiat’s lesson and design copyright to maximize copyright owner revenue, in the hope that it would maximize employment in the copyright industries, we will reduce welfare. To see this, consider the following simplified general equilibrium model. For the model, we will assume a continuum of potential products in a two-sector economy. Resources can be devoted either to the production of an original work of authorship in one sector of the economy or a non-copyrightable product in the other. The initial product in each sector generates a social value of one hundred, with the social value of each additional product decreasing in a linear fashion, on a one-for-one basis. Thus, the second product has a social value of ninety-nine, the third ninety-eight, and so on. We have one hundred units of a resource available that can be used to create either one product in the copyright sector or one product in the non-copyright sector, and all resources are fully employed. Thus, the sum of works in the copyright sector, \( n \), and the sum of products in the non-copyright sector, \( m \), will equal one hundred. In the absence of copyright, an individual who authors a new original work in the copyright sector will capture some fraction, \( \phi \), of that product’s social value. Similarly, an individ-

---

117. Consider agriculture as an example. Over the course of the twentieth century, the farm population fell from 32.5 million or 31.9 percent of the total U.S. population in 1916 to 2.9 million or only 1 percent of the U.S. population in 2006. JULIAN M. ALSTON ET AL., PERSISTENCE PAYS: U.S. AGRICULTURAL PRODUCTIVITY GROWTH AND THE BENEFITS FROM PUBLIC R&D SPENDING 9-11, 16 (2010). Yet, because of technological advances, the value added by this much smaller farm workforce grew nearly six-fold from 1929 to 2006, from $17 billion to $98 billion (measured in 2000 prices). Id. at 10. It’s true that these same technological advances cost the farm sector millions of jobs. Nevertheless, by releasing these workers into the rest of the economy, these technological advances, and their associated job losses, made possible the tremendous growth that occurred in the economy generally. U.S. Gross Domestic Product increased thirteen-fold from 1929 to 2006, from $866 billion to $11.3 trillion. Id. One can scarcely imagine a more foolhardy government policy than one that, in the name of protecting jobs, tried to limit these technological advances in order to keep one-third of our population employed on farms.
ual who creates a new product in the non-copyright sector will capture some fraction, $\phi_n$, of that product’s social value. The available resources are allocated between the two sectors so that the marginal work in the copyright sector and the marginal product in the non-copyright sector receive the same “price” or return on the resource; thus $\phi_c S_{cn} = \phi_s S_{nm}$, where $S_{cn}$ and $S_{nm}$ are the social value of the marginal work, $n$, and the marginal product, $m$, respectively. We will further assume that, in the absence of copyright, an individual captures only ten percent of the social value of a new original work, while an individual captures twenty percent of the social value of a new non-copyrightable product.¹¹⁸

Given this set-up, we can calculate the revenue captured by the copyright sector¹¹⁹ and the social welfare attributable to both sectors as we increase the level of copyright protection provided.¹²⁰ As copyright protection increases, it increases the fraction of the value that an author or copyright owner captures with respect to a new original work from ten percent to one hundred percent. Figure 6 presents the results. For welfare, I use the welfare obtained with no copyright as a baseline and present the welfare obtained at any given level of copyright as a gain or loss relative to the “no copyright” welfare. Welfare in the chart thus represents the gain in welfare or the loss in welfare for any given level of copyright protection, compared to a world without copyright.

¹¹⁸. While these fractions are assumptions, I have based them on my own sense for their likely values. An innovator or author will capture the full “value” of her work, and thus the fraction, $\phi$, will equal one only if: (i) the legal rules fully insulate the new product from competition; and (ii) perfect price discrimination can be achieved without cost. More plausibly, if the legal rules protect a new product from competition for its full economic life and the innovator sets a uniform profit-maximizing price for the product against a backdrop of a downward sloping, linear demand curve, and constant marginal costs, then the innovator will capture half of the new product’s social value, and the fraction will equal 0.5 or 50 percent. Of the remainder, it will be evenly split between consumer surplus and deadweight loss (representing unrealized social value). See Lunney, supra note 9, at 864-67; Lunney, supra note 24, at 557 n.283. I am estimating the fraction at twenty percent for non-copyrighted creative products based largely upon Mansfield, Schwartz, and Wagner’s work and their conclusion that most new products, whether patented or not, face competition within four years. Mansfield, Schwartz, & Wagner, supra note 60, at 914 & n.1 (“In the bulk of the cases, the new product could have been imitated in 2 years or less even if the imitator carried out the project at the most leisurely pace. In practically all cases it could be imitated in 3 years or less.”). I use ten percent as the fraction for copyrighted works to acknowledge that works of authorship can generally be copied more easily, more quickly, and less expensively than new products elsewhere in the economy. Even if we vary these assumptions, the essential insights of the model remain. Both the optimal level of copyright at which welfare is maximized and the crossover point at which copyright begins to reduce welfare are reached well before copyright owner revenue is maximized.

¹¹⁹. Rents or revenue to the copyright sector will equal $\phi_c^*(100n-0.5n^2)$.

¹²⁰. Total social welfare will equal $(100n-0.5n^2)+(100m-0.5m^2)$. 


Figure 6. The Effect of Increasing Copyright on Social Welfare and Copyright Owner Revenue.

Figure 6 illustrates several important points. First, point (1) represents the optimal level of copyright protection. This is the level of copyright protection that generates the highest social value. It does so by ensuring that the available resources are devoted to their highest valued use. With no copyright, resources are flowing elsewhere (into the non-copyright sector) when they could be used more valuably to create additional works of authorship. By correcting this, copyright can generate real welfare gains and ensure that resources flow towards the creation of additional works when that represents their most valuable use. Yet, as Figure 6 illustrates, only a very minimal copyright protection regime is necessary to achieve that goal. Once copyright becomes broader than necessary to equalize (approximately) the return between the copyright and non-copyright creative sectors, it begins to lead to the same misallocation of resources that it was meant to correct. Resources continue to flow into the creation of additional works when those resources would otherwise have been more valuably used elsewhere. As copyright continues to broaden, welfare losses begin to mount. While overly broad copyright leads to too many resources being devoted to original works, rather than too few, the welfare loss resulting from the resource misallocation is the same.  

Thus as copyright becomes broader and moves past point (1), welfare begins to fall.

This brings us to point (2): the crossover. As broader copyright draws more and more resources into the production of additional

121. See Patent Study, supra note 12, at 58-60; Robert M. Hurt & Robert M. Schuchman, The Economic Rationale of Copyright, 56 AM. ECON. ASS‘N 421, 429-30 (1966); Lunnery, supra note 24, at 599-601. While this seems counterintuitive, it follows from the same economic logic that justifies copyright. If we under-protect original works, we will have too few original works. “Too few” has a specific economic meaning. Specifically, it means that resources are flowing elsewhere when they would be more valuably used to create additional works. If we overprotect original works, we will have too many original works in the exact same sense. Resources will go towards creating additional works when those same resources would otherwise have been more valuably used elsewhere in the economy.
works, we quickly reach a point where the welfare losses from broader copyright are such that we would be better off with no copyright at all. In Figure 6, once we have past point (2), copyright has become so broad that we would be better off with no copyright at all. At this point, copyright is attracting too many resources into the creation of additional works and consequently leaving too few resources for the rest of the economy. The welfare losses resulting from this inefficient allocation of the available resources are just as bad as those created by having no copyright at all. Given the assumptions I made to generate Figure 6, we reach the crossover point in our model when the system of legal rights copyright creates would enable copyright owners to capture more than forty percent of the value of their original works. To be clear, I do not intend Figure 6 to define the precise point at which we reach this “no copyright is better” crossover in the real world. Rather, the point of Figure 6 is that such a crossover exists and we will find ourselves at the crossover point well before copyright becomes broad enough to enable copyright owners to capture a work’s full value and also well before copyright becomes broad enough to maximize revenue to copyright owners.

This brings us to point (3), the point at which copyright maximizes revenue to copyright owners. As Figure 6 reflects, revenues to the copyright industries continue to increase over the full range of potential copyright protection. As broader copyright enables a copyright owner to capture an ever-greater fraction of the value associated with her work, revenue to the copyright industries steadily increases. This increased revenue comes from increasing the revenue associated with: (i) with preexisting works; and (ii) with the additional works that broader copyright brings forth. Thus, if the goal is to maximize copyright industry revenue, with the hope thereby of maximizing copyright sector employment, then the appropriate response is to design copyright so that authors capture the full value of their respective works.\(^{123}\)

\(^{122}\) Whether it does so depends on how accurately its assumptions match the real world markets against which the production of original works competes for resources. While I find the assumptions to be a reasonably good match for real world markets, others might plausibly disagree. We should also note that the welfare calculations in Figure 6 represent only those welfare gains and losses attributable to resource allocation. I have omitted any consideration of the overprotection costs that arise from copyright’s uniformity. Had we included those costs as well, the welfare gain associated with any given level of copyright would have been reduced and the welfare losses associated with any given level of copyright would have been larger. In addition, the model assumes a uniform distribution in the additional works that broader copyright brings forth. If, as is more likely, the additional works exhibit a normal distribution as shown in Figure 1, the welfare losses from broader copyright would be higher.

\(^{123}\) In the real world, ever-broader protection may not maximize copyright industry revenue. As Jim Bessen and Mike Meurer have argued in the patent context, if copyright protection becomes sufficiently broad, then it may become difficult to introduce new works as they will face too many claims of copyright infringement from earlier works. JAMES
As Figure 6 makes clear, however, designing copyright to maximize copyright industry revenue and employment does not increase welfare; it decreases welfare. In welfare terms, the additional revenue to the copyright industries does not represent a net gain for society; it represents mere wealth redistribution from consumers to copyright owners. Every additional dollar that flows to the copyright industries must be taken from consumers (or perhaps, as Bastiat argued, from other industries). Indeed, given the inevitable administrative and transaction costs broader copyright entails, to put one dollar into the hands of copyright owners, we have to take somewhat more than a dollar from consumers. This redistribution does not therefore represent a welfare gain; instead it represents a welfare loss.

It may be that in pursuing the goal of maximizing copyright owner revenue, copyright has already become so broad that it is past and perhaps well past point (2), the crossover point. Yet, whether we have already reached the crossover point or not, Figure 6 firmly re-
futes the notion that attempting to maximize revenue to and employment in the copyright sector represents a sensible policy goal. Going beyond point (1), which represents the efficient level of protection, may increase revenue to and jobs in the copyright industries, but it will take that revenue and those jobs from elsewhere in the economy. Moreover, once we are past point (1), additional jobs in the copyright sector will generate less value than the jobs taken away elsewhere in the economy. In short, designing copyright to maximize copyright owner revenue will only make us worse off.

C. Market Failure in Product Markets and Political Markets

Unfortunately, there is good reason to believe that copyright today is already well past the crossover point and is affirmatively making us worse off than we would be with no copyright. That reason is Congress. Just as markets for goods and services will fail to achieve a Pareto optimal allocation of resources in predictable circumstances, so too will political markets. In a representative democracy, political markets fail, *inter alia*, when the interests of a dispersed interest group, such as consumers, and those of a concentrated interest group, such as the copyright industries,\textsuperscript{125} collide. Whether we think of our elected representatives as public-minded servants of the people or as venal and corrupt crooks merely waiting their turn to be indicted, Congress is systematically likely to enact measures that benefit a concentrated group at the expense of a dispersed group, even when the measures represent nothing more than undesirable rent-seeking.

This political failure arises because a concentrated group has a significant collective action advantage over a dispersed group in organizing to influence the course of proposed legislation. For any given amount at stake for the group as a whole, the members of a dispersed group will have less individually at stake than will the members of a concentrated group. Given that becoming involved in a political fight entails information and transaction costs, as an individual’s personal stake becomes increasingly small, it becomes increasingly rational to remain uninformed and uninvolved. Moreover, because legislation is a public good, as a group gets larger and more diverse, individuals

become more likely to rely (or free ride) on the efforts of others.\(^{126}\) In contrast, when proposed legislation would impact a smaller and more cohesive group,\(^{127}\) each member of the group has a larger individual stake in the fight and hence is more likely to become informed and involved. Moreover, a smaller group is more likely to have effective mechanisms, such as trade associations, to coerce participation by each member and to thereby limit free riding.\(^{128}\)

For these reasons, a concentrated group will have more resources available, for a given dollar amount at risk, than will a dispersed group.\(^{129}\) In addition, the concentrated group is likely to prove more effective in using the resources available to influence the legislature than a dispersed group.\(^{130}\) An honest politician, lacking perfect infor-

\(^{126}\) In other words, political markets for legislation are likely to fail for precisely the same reason that ordinary markets fail for public goods.

\(^{127}\) Whether a group is concentrated or dispersed depends more on its “effective size” than the absolute number of affected individuals. Effective size refers to the fact that corporations, unions, and other well-organized groups count as only one “effective person,” even though the corporation may have thousands of employees and shareholders. In addition, “effective” delineation recognizes that, while the total costs of a measure may be imposed on a very large group, the primary cost may fall on a select subgroup. For example, air pollution control might impose ninety-five percent of its cost on fifty corporations and the remaining five percent on five million individuals with wood-burning fireplaces. Determining whether an industry is concentrated presents the same problem: five major corporations may control ninety-five percent of the business with 500 other corporations controlling the remaining five percent. To determine concentration in a market, a court can look to the market share of the top four to eight corporations, see George J. Stigler, Free Riders and Collective Action: An Appendix to Theories of Economic Regulation, 5 Bell J. Econ. & Mgmt. Sci. 359, 362 (1974), or use the Herfindahl-Hirschman Index, see generally Neil B. Cohen & Charles A. Sullivan, The Herfindahl-Hirschman Index and the New Antitrust Merger Guidelines: Concentrating on Concentration, 62 Tex. L. Rev. 453 (1983). Using such an approach we can readily see that although broadening copyright implicates copyright interests owned by millions of people—since everyone who has ever written a personal e-mail is a copyright owner—the effective size of the interest group benefitted by broader copyright has, historically, been much smaller than these large numbers suggest because most of the benefit falls on the handful of movie studios, publishing houses, and record labels.


\(^{129}\) See Lunney, supra note 125, at 1951 (“For a given dollar amount at risk, each member of a concentrated group faces a higher individual risk than would a member of a dispersed group. In addition, a concentrated group can more easily coerce each of its members to participate than can a dispersed group. As a result, each member of a concentrated group is more likely to participate in the lobbying effort and less likely to free ride on the efforts of others.”).

\(^{130}\) See Lunney, supra note 125, at 1950-51 (“[A concentrated] group can coordinate and control its members’ efforts, achieving economies of scale. The group can avoid duplication of effort, obtain expert support for its position, and even hire a full-time lobbyist to protect its interests. When the legislature requires information on costs or technology within the exclusive control of the members of an interest group, the concentrated group can
mation with respect to the desirability of proposed legislation, may use the support and opposition that she hears as evidence of the measure’s desirability. But in a fight over broadening copyright, such a politician is likely to hear strong support for and little opposition to such a measure. This apparent support is not, however, because the measure is desirable; rather, it derives from the huge collective action advantage the copyright industries enjoy over consumers generally. For venal politicians, that same collective action advantage translates into disproportionate resources for copyright owners to buy such politicians and their votes.

As a result, Congress is systematically likely to enact legislative measures that broaden copyright even when they reduce welfare.131 As Figure 6 shows, even when broadening copyright would generate welfare losses, it continues to increase copyright owner revenue. As self-interested actors, copyright owners care about their revenues, not about social welfare. We should therefore fully expect copyright owners to go to Congress and fight for broader copyright and the associated increase in their revenue, even if broader copyright generates welfare losses to society as a whole. Given the market concentration in the various copyright industries,132 we should expect that lobbying effort to be well-funded and effective, which it is. Of course, every additional dollar in revenue to copyright owners means a dollar (and then some) lost by consumers. But we should not expect consumers to prove effective, generally, in lobbying against broader copyright. Even when broader copyright would transfer hundreds of millions, even billions, of dollars from consumers to copyright owners, each individual consumer has only a few dollars at stake if a given proposal to broaden copyright passes. That individual cost will likely prove to be insufficient and, indeed, has proven insufficient historically to persuade many consumers to become personally involved in lobbying against broader copyright.

The net result of this collective action advantage, consistently pressed over the last two hundred years, has been a steady expansion of copyright’s term and scope. Most of this expansion has been undesirable; for the most part, it has represented mere rent-seeking: transferring wealth from consumers to copyright owners with no wel-

131. See, e.g., Lunney, supra note 9, at 895-901.

132. See, e.g., Lydia Pallas Loren, Untangling the Web of Music Copyrights, 53 CASE W. RES. L. REV. 673, 719 (2003) (noting that “in the music industry there is the very real problem of industry concentration”); Lunney, supra note 9, at 916 & n.325 (citing sources showing that in the movie industry in the year 2000, 70.4% of domestic box office revenue went to six studios and that the five record labels in the Napster case had a market share of roughly 75%).
fare gains. Indeed, much of the expansion has likely generated considerable welfare losses. While undesirable, this expansion has come because the collective action advantages of copyright owners have tricked Congress (or Congress has allowed itself to be tricked)\textsuperscript{133} into believing that the expansion was desirable. The opposition to PIPA and SOPA was successful because it enabled consumers, at least for one brief moment, to let Congress know how they truly felt about broader copyright. The online petition drive both gave consumers the information that they needed to determine whether they wanted to oppose the measures and provided a means to express their opposition. Moreover, it provided consumers with both at a very low cost—a cost sufficiently low that it was rational for individual consumers to take the time to become involved.

Whether the forces that defeated PIPA and SOPA can be converted into a political movement that will accomplish true copyright reform in the United States remains to be seen. While I am not terribly optimistic, I am hopeful. Canada recently managed several significant steps in the right direction.\textsuperscript{134} Given the flaws in the political process and the resulting likelihood of political failure in this area, it is too bad, really, that Thomas Jefferson could not persuade James Madison to limit Congress’s constitutional power in this area to a specific term of years.\textsuperscript{135} Had he done so, copyright would be far less overbroad, at least, in terms of duration, than it is today.

\textsuperscript{133}. While Congress likely includes a mixture of honest and venal politicians, actions that seek to limit the presentation of competing viewpoints or access to other objective sources of information regarding a proposed measure’s desirability, such as Representative Smith’s decision to limit witnesses before proceeding to the mark-up of SOPA and Congress’s earlier decision to de-fund the Office of Technology Assessment, \textit{cf.} Pamela Samuelson, \textit{Book Note, Is Copyright Reform Possible?}, \textit{126 HARV. L. REV.} 740 (2013) (recommending the re-establishment of the OTA), tend to support the view that our elected representatives are predominantly venal.

\textsuperscript{134}. At the judicial level, the Canadian Supreme Court has recognized a broader approach to fair dealing in several recent cases and has also reiterated its position that copyright exceptions such as fair dealing should be treated as users’ rights. \textit{See Soc’y of Composers, Authors and Music Publishers of Can. v. Bell Can.}, [2012] S.C.R. 326, 327 (Can.) (holding that providing thirty-second previews of music to consumers constitutes fair dealing); \textit{Alta. (Educ.) v. Canadian Copyright Licensing Agency (Access Copyright)}, [2012] S.C.R. 345, 349-52 (Can.) (holding that the copying of short excerpts from complementary texts by teachers constitutes fair dealing). In its recent statutory reform, Canada also limited the availability of statutory damages for non-commercial purposes. Copyright Modernization Act, S.C. 2012, c. 20, cl. 46 (Can.) (limiting statutory damages in cases involving infringement(s) for non-commercial purposes to a maximum of $5,000 (Canadian) for all infringements in a single proceeding for all works).

\textsuperscript{135}. In response to James Madison’s proposal to limit Congress’s power in this area “for limited times,” Jefferson responded:

\begin{quote}
I like it as far as it goes; but I should have been for going further. For instance the following alterations and additions would have pleased me. . . . Art. 9. Monopolies may be allowed to persons for their own productions in literature and
\end{quote}
Perhaps rather than copyright reform, it is time for our own Statute of Monopolies moment. In 1624, the English Parliament rose up and rejected the English Crown’s practice of bestowing the exclusive right to manufacture, import, or sell various commodities on one court favorite or another. As with copyright, royal patents in the 16th and 17th centuries initially served a public purpose and were granted to help ensure the introduction of new trades into Britain. Over time, however, they became undesirable monopolies granted “as rewards for political patronage.” To restrain the Crown’s power and to ensure that it served a public purpose, the English Parliament enacted the Statute of Monopolies.

Perhaps, it is time for “We the People” to rise up and deprive Congress of the power to provide exclusive rights in works of authorship. Thomas Jefferson believed that we could and would take such action should it prove necessary. As he wrote to Madison in discussing the need for limits on Congress’s power to grant exclusive rights to authors and inventors: “These restrictions I think are so guarded as to hinder evil only. However if we do not have them now, I have so much confidence in my countrymen as to be satisfied that we shall have them as soon as the degeneracy of our government shall render them necessary.”

The degeneracy of our government on the issue of copyright has become increasingly apparent. The only real question at this point is whether we will prove ourselves worthy of Jefferson’s faith.

While I recognize the political difficulty and perhaps futility of proposing a constitutional amendment limiting Congress’s power in this area, I think it is both the time and past the time to put such options on the table. It has been over two hundred years since our


137. Adam Mossoff, Rethinking the Development of Patents: An Intellectual History, 1550-1800, 52 HASTINGS L.J. 1255, 1265 (2001); see also Thompson v. Haight, 23 F. Cas. 1040, 1042-43 (C.C.S.D.N.Y. 1826) (No. 13,957) (Van Ness, District Judge) (“[T]hese pernicious expedients for increasing the revenue, or replenishing the exhausted coffers of the crown, were never employed in the extent to which they were pushed by [Elizabeth]. Elizabeth lavished them, with a munificent hand, upon her courtiers and her servants, whether distinguished by her personal favour or for their public services. All trade and commerce, whether foreign or domestic, was appropriated by monopolists.”).

138. See Mossoff, supra note 137, at 1270-73; see also Ochoa & Rose, supra note 136, at 913 (discussing the Statute of Monopolies and its exceptions).

139. Letter from Jefferson to Madison, supra note 136, at 368.
Constitution was written, and we have a much better sense today for where representative democracy works and where it fails. Because copyright benefits a concentrated and well-organized interest group at the expense of a dispersed group, establishing an optimal copyright regime is simply not something Congress has done or will do well. We should therefore limit Congress’s power to act on this issue. At the simplest, such a constitutional amendment might follow Jefferson’s suggestion and substitute “for no more than fourteen years” for the phrase “for limited times” in Article I, section 8, clause 8. Taking it a step further, an amendment might specify or limit the nature of the “exclusive rights” that Congress may grant. I fully recognize that such an approach would enshrine a set of rights that, even if optimal today, may not prove optimal for all time. Such an approach would almost certainly impose a set of legal rights that will not perfectly fit the needs of the future, as technology and markets change. Nevertheless, I believe that such an approach remains preferable to our current approach. Any welfare losses that may result from constitutionalizing today’s optimal set of rights and imposing those rights onto the future would be less than the welfare losses that will result and have resulted from leaving the issue to Congress. Given how overbroad copyright has become, even an amendment barring Congress (and the states as well) from granting exclusive rights to authors for their writings altogether would likely be better than where we find ourselves today.

In the end, the question is not whether the market for original works will, if left to itself, achieve a Pareto optimal allocation of resources. Very few markets work perfectly. If we had no copyright and essentially left the market to its own devices, there will undoubtedly be some instances where resources flow to other uses even though creating an additional original work represents the most valuable use of those resources. Even so, the real question is not whether some degree of market failure will otherwise occur, but whether Congress can improve on the market’s outcome, given the likelihood of political failure on this issue. Given how well music output has held up in the face of widespread file sharing and given the excesses of today’s copyright law, we have far more to fear from congressional action than we do from leaving the market for original works alone.

V. CONCLUSION

As a test case of our brave new digital world, the music industry’s experience suggests that we can have both widespread consumer copying and sufficient incentives to produce new original work. Rather than decrease in the face of widespread consumer copying, output in the music industry, both in terms of quantity and quality, has increased and by some measures has increased sharply. If copyright’s
purposes is indeed a public one, “to promote the Progress of Science,” then the fact of increased copyright output would seem to fully rebut any need for increased copyright protection.

Yet copyright owners insist that they need more copyright. Rather than argue that widespread consumer copying is leading to fewer original works—because it is not—copyright owners have argued instead that more protection means more revenue for copyright owners and hence more jobs in the copyright sector. 140 Even if true, this argument provides no basis for legislative action. As Bastiat long ago explained, government action that increases revenue to one sector of the economy necessarily reduces revenue to every other sector of the economy. Whether this is done through a direct tax-and-subsidy scheme or through an indirect subsidy such as copyright, the result is the same. If broader copyright means more revenue and more jobs in the copyright sector, it necessarily means less revenue and hence fewer jobs everywhere else.

For that reason, the jobs argument cannot justify broader copyright. While the public good character of original works and the ease with which they can otherwise be copied may justify some level of copyright protection, they justify only that level of protection needed to enable individuals to recover a fraction of their respective work’s value both comparable to and neither higher nor lower than that recoverable for creative work elsewhere in the economy. As a practical matter, ensuring such equality requires only a very limited degree of copyright protection—something akin to the fourteen years of protection against mechanical duplication by competing commercial publishers that the 1790 Copyright Act provided. 141

The widespread consumer copying that digital technology and the Internet have made possible are a recent phenomenon. While we still do not know all of the ways in which digital technology will impact the creation and distribution of new original works, both economic theory and empirical data to date suggest that consumer copying, even if widespread, does not pose a threat to copyright’s constitutional delimitation of promoting the “Progress of Science.” Consumer copying, including unauthorized file sharing, increases access to and the value of original works directly. 142 While it may reduce copyright

140. See supra notes 25-32 and accompanying text.

141. While the 1790 Copyright Act included a fourteen-year renewal, in addition to the fourteen-year primary term, not many works were renewed. See LAWRENCE LESSIG, FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY 319 nn.9-10 (2004) (“Few copyright holders ever chose to renew their copyrights. For instance, of the 25,006 copyrights registered in 1883, only 894 were renewed in 1910.”).

owner revenue, the music industry’s experience suggests that, even with widespread consumer copying, we will still have incentives that are sufficient and, perhaps more than sufficient, to ensure the creation and dissemination of new original works.

To be sure, it may turn out that the music industry’s experience will not prove representative. In other fields, in the face of widespread consumer copying, we may see actual reductions in creative output. But it will be soon enough to act when creative output actually falls. That consumer copying has led to falling revenue and perhaps reduced employment for particular copyright industries does not and cannot justify expanding copyright. If we pretend otherwise and, in doing so, increase copyright protection in an attempt to maximize copyright owner revenue and, perhaps, copyright industry employment, we will only be left worse off.

THE COST OF AVOIDANCE: PLURALISM, NEUTRALITY, AND THE FOUNDATIONS OF MODERN LEGAL ETHICS

MELISSA MORTAZAVI

ABSTRACT

This Article offers an answer to key questions in modern American legal ethics: when and why did the legal profession stop talking about professional conduct in moral terms? Mining the history of current rules governing lawyer conduct, this Article reveals that while the 1969 Model Code of Professional Responsibility sought to revolutionize legal ethics by creating a professional code that was more transparent, democratized, and less hierarchical than the preceding 1908 Canons of Legal Ethics, that effort also excised a moral understanding of lawyering in order to facilitate a particular understanding of pluralism.

The drafters of the 1969 Model Code faced a difficult task. They recognized women and minorities were entering the legal profession in unprecedented numbers. Aware of impending conflicts within the newly diverse bar and unsure how to resolve them, drafters of the Model Code struck a devil's bargain: in exchange for the peaceable coexistence of heterogeneous parties, the Model Code sought to remove moral disputes from the workplace by embracing neutral partisanship. This Article discusses the consequences of that choice. It argues that in order to permit one form of pluralism (demographic pluralism) the bar adopted a professional conduct system (neutral partisanship) that now impedes the inclusion of full substantive pluralism (including value pluralism).

Neutrality is not neutral. Avoidance has its costs. The Model Code did not actually remove morality from practice: it only prevented new lawyers from having the language and means to challenge and change existing moral norms in the profession. Modern legal ethics’ endorsement of neutral partisanship structurally impedes substantive discussions amongst students, lawyers, judges, and academics about proper ends and appropriate means. This Article is a call to reopen discussion as it reveals why the legal profession embraced this particular model of lawyering in the first place and how that purpose has been frustrated over time.

I. INTRODUCTION

II. THE CANONS OF ETHICS AND THE MODEL CODE OF PROFESSIONAL RESPONSIBILITY

A. In the Beginning, the ABA Created the Canons

B. From Canons to Code: Operationalizing Theory

III. CODIFICATION AND STRUCTURE

A. Rules Adopted; Ethics Orphaned

B. Enforceability: The Model Code and Discipline

IV. STANDARDIZING THE STANDARD CONCEPTION

A. Neutral Partisanship: A Primer

B. A Modern Moral Justification of Neutral Partisanship

C. Neutral Partisanship as a Vehicle for Client Autonomy

1. Candor, Loyalty, and Effective Client Service

2. Codifying Trust: The Model Code’s Bind

3. Clients and the Knowledge Deficit

* Visiting Assistant Professor of Law, Brooklyn Law School. I would like to thank Anita Bernstein, W. Bradley Wendel, Nelson Tebbe, Russell Pearce, Bruce Green, Rebecca Roiphe, Stephen Lee, Alan Trammell, Emily Berman, Robin Effron, Sarah Cravens, and the faculty workshops at Berkeley Law, University of Arkansas School of Law, University of Oklahoma College of Law, and Brooklyn Law School for their comments on earlier drafts of this piece. Most of all, my deepest gratitude to Roger Michalski.
I. INTRODUCTION

By the 1960s, the American Bar Association (ABA) could not remain impervious to the general social upheaval of the civil rights revolution. The legal profession was in flux and faced changes in its overall size and composition as a wave of new entrants from formerly excluded groups gained broader access to higher education and political capital. Lawyers and legal regulators faced novel and vexing questions: How could the bar devise ethical standards to include these newcomers, appeal to current members, and distance itself from...
its exclusionary past? What moral views regarding the role of lawyers and hot-button topics like civil disobedience would female or minority attorneys hold? If they did have different views than the existing bar, “whose conscience and whose ethical standards [we]re to control”

The 1969 Model Code of Professional Responsibility (Model Code) was the bar’s answer to these questions. In it, the bar laid out professional principles to govern lawyers in an “urbanized society” and break with the dated and elitist 1908 Canons of Ethics. This was no easy task. The mechanics of a new, more inclusive system of professional conduct were not obvious. Faced with moral and demographic pluralism at the bar and unwilling or unable to negotiate it, the drafters of the Model Code struck a devil’s bargain; in exchange for


5. See infra notes 85-91 and accompanying text. Much of the established bar viewed civil disobedience skeptically, while newer entrants may have supported opposition of unjust laws through civil disobedience. In the view of the existing bar, lawyers had no role in fomenting disobedience of established laws, even if these laws were oppressive, for lawyers instead ought to work through established legal channels to effectuate change. See, for example, 111 CONG. REC. 15103 (1965), in which one of the leading legal ethics reformers—in an address originally given to the Tennessee Bar Association on June 17, 1965, introduced into the record by Strom Thurmond—warned of civil disobedience’s ability to “seriously threaten the breakdown of law, order, and morality” and called for “impartial, even-handed, vigorous, swift and certain enforcement of our criminal laws, and the real and substantial punishment thereunder of all conduct that violates those laws.” See also id. (arguing that “[n]o ‘end’ . . . however worthy [can ever] justify resort to unlawful means” and that “America needs a genuine revival of respect for law and orderly process . . . a new impatience with those who violate and circumvent the laws, and a determined insistence that laws be enforced” (alterations in original)).


7. See Geoffrey C. Hazard, Jr., The Future of Legal Ethics, 100 YALE L.J. 1239, 1259 (1991) (“[R]adical changes occurring in the profession weakened the traditional bar’s conception of itself, which in turn enhanced the bar’s difficulties in dealing with the fact that its norms were becoming public law.”); Wolfram, supra note 2, at 210 (“[S]trresses within our society . . . affected American law and the American legal profession. These stresses created a fertile ground for legal change to occur and appear to have combined as its triggering force.”). See generally Chris G. McDonough & Michael L. Epstein, Regulating Attorney Conduct: Specific Statutory Schemes v. General Regulatory Guidelines, 11 Touro L. REV. 609, 610-11 (1995); Fred C. Zacharias, Foreword: The Quest for a Perfect Code, 11 Geo. J. LEGAL ETHICS 787 (1998).

8. MODEL CODE OF PROF’L RESPONSIBILITY Preface (1980); see also Edward L. Wright, Study of the Canons of Professional Ethics, 11 CATH. LAW 323, 323 (1965) (“[T]he Canons of Professional Ethics of the American Bar Association need revision because . . . changing conditions in an urbanized society require new statements of professional responsibility.”).

the peaceable coexistence of heterogeneous parties, they excised discussions of morality from the workplace.

This Article discusses the consequences of that choice. It argues that in order to facilitate one form of pluralism (demographic pluralism) the bar adopted a professional conduct system (neutral partisanship) that now impedes the inclusion of full substantive pluralism (including value pluralism). It did so by creating a set of national disciplinary rules that removed discussions of morality from professional discourse. Those rules incorporated, strengthened, and operationalized previous loose commitments to a client-centric model of lawyering now known as “neutral partisanship.” Just as color-blindness became the new hegemonic paradigm elsewhere, the legal profession adopted its own sanitized regime for regulating a diverse bar and avoiding conflict through the neutral-partisan model.

However, neutrality is not neutral. Avoidance has its costs. Modern legal ethics’ endorsement of neutral partisanship structurally impedes meaningful discussions amongst students, lawyers, judges, and academics about proper ends and appropriate means. The Model Code did not succeed in removing morality from standards of practice; it only prevented new lawyers from developing the language and means to challenge and modify existing moral norms in place. The Model Code set a certain moral vision of legal professionalism: lawyers were not expected to be any more than legal technicians, fulfilling clients’ ends. Certain skills were prioritized, others cast aside. A lawyer’s moral contemplation, empowerment, or even responsibility was rendered officially optional. As a result, today’s legal profes-

10. See, e.g., Edward L. Wright, The Code of Professional Responsibility: Its History and Objectives, 24 ARK. L. REV. 1, 10 (1970) (chairman of the drafting committee noting that “[t]he division [between issues concerning morals versus issues requiring disciplinary action] is one that has not previously been generally made” in legal ethics and that “[o]ne of the weaknesses of the Canons of Professional Ethics was its failure to speak to the two forces separately”); Singleton v. Stegall, 580 So.2d 1242, 1244 n.4 (Miss. 1991) (“Our rules regulating professional conduct have evolved from canons to ethical considerations and now to a code quite like unto a criminal code.”).

11. Neutral partisanship is the idea that lawyers are not accountable for the morality of their client’s chosen ends (hence neutrality) and yet act as a partisan in favor of their client’s interests (by providing representation that argues on their client’s behalf). This concept has many attendant names including “amoral lawyering” or “role morality” and “role differentiation,” since the role of being a lawyer is viewed as morally distinct from actions taken on from the “role” of being a private citizen. See Stephen L. Pepper, The Lawyer’s Amoral Ethical Role: A Defense, a Problem, and Some Possibilities, 1986 AM. B. FOUND. RES. J. 613 (using the term amoral lawyering model due to the distance of the lawyer’s moral accountability for client outcomes); Norman W. Spaulding, The Rule of Law in Action: A Defense of Adversary System Values, 93 CORNELL L. REV. 1377, 1378 (defining role morality as “the idea, much maligned by legal ethicists, that lawyers should receive some degree of immunity from the general requirements of conscience on account of their distinctive social role”).
sion suffers from a self-imposed inability to talk about morality or fundamental differences in the workplace.

Equally troubling, this neutral-partisan ideal renders all lawyers interchangeable widgets, as the bulk of what makes lawyers individuals is deemed professionally irrelevant, worthless, and even inappropriate. Neutral partisanship is the dominant moral fiction where all people are expected to behave the same way, and that singular way of acting is deemed “neutral”—although, in reality, it is not.

This Article tells the story of how and why neutral partisanship became the dominant norm for modern lawyering and argues that while such “neutrality” may have at one time served the purpose supporting pluralism at the bar, today it has outlived its utility. Part II begins by laying out how the bar came to pursue demographic pluralism in drafting the Model Code and why the adoption of neutral partisanship became a vital part of that process. Part III discusses how the Model Code set up a system that solidified a commitment to neutral partisanship by divorcing morality from mandatory rules governing professional conduct. Finally, Part IV explains how neutral partisanship actually impedes a full understanding of pluralism which includes value pluralism. It does so by critiquing 1) the argument that neutral partisanship placed the power in the hands of the clients (the people) over “elitist” lawyers and 2) the idea that neutral partisanship itself is (or can be) value neutral.

I conclude that the Model Code has failed to serve the very values that it sought to facilitate: broad conceptions of pluralism. As such, the ongoing disenfranchisement of discussions of morality in modern professional discourse cannot be justified on those terms. However, by recognizing the historically contingent normative commitments embedded in modern legal ethics, the bar and the academy can work towards revitalizing legal ethics so that it becomes a tool for the modern bar, not a hindrance to it.

12. Current norms, rules, and regulations governing lawyers today exist because of the shift in legal ethics from a treatise-style discussion of professionalism to a statute-like structure that polices the conduct of lawyers—a shift that occurred with the advent of the Model Code. See, e.g., Developments in the Law—Lawyers’ Responsibilities and Lawyers’ Responses, 107 HARV. L. REV. 1547, 1582 (1994) (“Unlike the Canons, the Model Code provided specific and legally binding normative rules, thus ‘legalizing’ substantive professional regulation.”); Hazard, supra note 7, at 1251 (“In retrospect, it is clear that the crucial step in the ‘legalization’ process occurred in the change from the 1908 Canons to the 1970 Code, rather than from the Code to the 1983 Rules. It was the Code that first embraced legally binding norms in the form of the Disciplinary Rules . . . .”).
II. THE CANONS OF ETHICS AND THE MODEL CODE OF PROFESSIONAL RESPONSIBILITY

A. In the Beginning, the ABA Created the Canons

In the United States, the history of attempts to formalize legal ethics into a national document is concise. In part in response to the American Medical Association’s adoption of a professional code of ethics, as well as concern over excessive “commercialism” in law practice, the ABA created an advisory ethics committee in 1905. The Committee’s work culminated in a draft ethics proposal entitled the “Canons of Professional Ethics” (Canons). George Sharswood’s 1854 essay, “An Essay on Professional Ethics,” was of particular importance as it was reprinted and circulated to all ABA members with the draft Canons. The ABA membership ultimately adopted these Canons officially in 1908.

The Canons were brief, containing only thirty-two guidelines, and were written in broad terms. They read more like a treatise or essay

13. This is not to say, however, that no rules governed lawyer behavior prior to this time. Rather, the common law, in connection with norms of practice and homogeneity in the group of people trained in law, provided guidance on improper conduct.

14. Commercialism was code for opposition to the emergence of working class lawyers who served immigrants, the urban poor, and blue-collar workers. JEROLD S. AUERBACH, UNEQUAL JUSTICE: LAWYERS AND SOCIAL CHANGE IN MODERN AMERICA 41-43 (1976).

15. At the turn of the century membership in the ABA was by invitation only and limited to a small but highly influential percentage of the overall lawyer population. See JAMES E. MOLITERO, THE AMERICAN LEGAL PROFESSION IN CRISIS: RESISTANCE AND RESPONSES TO CHANGE 10 (2013).

16. Susan D. Carle, Lawyers’ Duty to Do Justice: A New Look at the History of the 1908 Canons, 24 LAW & SOC. INQUIRY 1, 7-8 (1999) (“[L]aw journals at the turn of the century were replete with articles lamenting growing ‘commercialism’ in law practice . . . .”).

17. David R. Papke, The Legal Profession and Its Ethical Responsibilities: A History, in ETHICS AND THE LEGAL PROFESSION 29, 37 (Michael Davis & Frederick A. Elliston eds., 1986). The Canons borrowed doctrinally from the ethics rules of Alabama, as well as the works of George Sharswood and David Hoffman upon which the Alabama Code of Ethics was based. THOMAS D. MORGAN & RONALD D. ROTUNDA, PROFESSIONAL RESPONSIBILITY: PROBLEMS AND MATERIALS 11 (7th ed. 2000) (discussing how Professor Sharswood’s work was influenced by David Hoffman’s 1836 publication Fifty Resolutions in Regard to Professional Development); Carle, supra note 16, at 9 (relaying how the drafters of the Canons consulted Hoffman and Sharswood in addition to existing state codes, particularly that of Alabama); Allison Marston, Guiding the Profession: The 1887 Code of Ethics of the Alabama State Bar Association, 49 ALA. L. REV. 471, 507 (1998) (discussing the role the Alabama Code served as a model for the original Canons).


than a rulebook.\textsuperscript{20} The Canons’ lofty tone was often distinctly moral in nature, concluding that a lawyer finds his “highest honor in a deserved reputation for fidelity to private trust and to public duty, as an honest man and a patriotic citizen,” not exclusively in the service of a client.\textsuperscript{21}

The level of generality of the Canons’ wording did not prevent it from becoming canonical quickly, in the sense of being widely adopted and influential.\textsuperscript{22} Although the ABA had (and has) no authority over the various state bars, over time the Canons were formally or informally adopted in most states through either direct bar or court action.\textsuperscript{23} Over the next sixty years, the ABA added only fifteen provisions to the Canons.\textsuperscript{24} Thus, the original 1908 Canons remained largely intact and central to governing legal practice in the United States well into the early 1970s.\textsuperscript{25}

Critiques of the Canons were strong and mounting by the 1960s.\textsuperscript{26} In 1958, the ABA Joint Conference had submitted an ethics report penned in large part by Lon Fuller, a celebrated ethics scholar, flagging the need to revisit the Canons. He noted that “[t]oday the lawyer plays a changing and increasingly varied role. In many developing fields the precise contribution of the legal profession is as yet undefined.”\textsuperscript{27} For some, concerns focused on the lack of specificity and en-

\begin{itemize}
\item \textsuperscript{20} See generally Final Report, supra note 18, at 575-85.
\item \textsuperscript{21} Id. at 584.
\item \textsuperscript{22} See Henry S. Drinker, Legal Ethics 23-26 (1953); Restatement (Third) of Law Governing Lawyers § 1, cmt. b (2000) (noting how states differed in treating the Canons as mandatory rules or nonbinding guidance for proper conduct); Papke, supra note 17, at 39 (noting how three-fourths of all states had adopted the Canons by the beginning of World War I).
\item \textsuperscript{23} But note, bar adoption is not necessarily coterminous with a consensus of support from practicing lawyers. Some scholars have noted that participation in bar associations is partially the purview of economic privilege, as time spent on bar associations necessarily cuts into time spent on billable or paying matters. See Papke, supra note 17, at 36 (“Country lawyers could rise to prominence in the bar associations of rural states, but in general urban lawyers, with the resources and types of practices that could facilitate conventional and organized bar work, were the leaders of the bar associations.”).
\item \textsuperscript{25} See Stephen Gillers, Regulation of Lawyers: Problems of Law & Ethics 11 (9th ed. 2012).
\item \textsuperscript{27} Id. at 1159.
\end{itemize}
forceability in the Canons. In keeping with the distributive justice sentiments of the time, others critiqued the profession’s failure to serve underrepresented groups of the American population. Some highlighted the darker undertones of the Canons’ naissance, arguing that it functioned as a gatekeeper to keep ethnic minorities and women from gaining upward mobility as lawyers and serving low-income and immigrant communities. The Canons were also susceptible to critique as maintaining the landed gentry’s exclusivity of the bar. With general, ambiguous language and limitations on advertising, fees, and client development, the Canons were a natural poster child for the established bar’s elitism and distance from less well-connected lawyers and clients.

These internal and external concerns finally came to a head in 1964. That year, the ABA president, and future Justice, Lewis F. Powell Jr., convened a “Special Committee on the Evaluation of Ethical Standards” (known as the “Wright Committee”) to propose amendments to the existing Canons. Originally, the Wright Committee convened with the purpose of recommending revisions to the

28. See Model Code of Prof’l Responsibility Preface (1980) (listing among the shortcomings to the Canons that “most of the Canons did not lend themselves to practical sanction for violations”); 58 Ann. Rep. A.B.A. 94-95 (1935) (arguing that the Canons offer little concrete guidance and suggesting “a Code of Practice which will deal not with general principles but with the specific abuses involved”); John F. Sutton, Jr., Guidelines to Professional Responsibility, 39 Tex. L. Rev. 391, 422-23 (1961) (arguing that the Canons are insufficiently specific to set a reasonable minimum standard); Wright, supra note 10, at 4 (quoting a 1958 American Bar Foundation report stating that the Canons do not present “sufficient detail” in dealing with “specific situations encountered in actual practice”).

29. See Papke, supra note 17, at 38, 41. See generally Wolfram supra note 4, at 485 (“[T]he ABA until well into the twentieth century functioned mainly as an exclusive social fraternal organization of high-status lawyers rather than as a broadly representative and unofficial regulatory body.”).


31. See Hazard, supra note 7, at 1250 (“The Canons . . . expressed the viewpoint of an economically advantaged social stratum distinguished by its intellectual accomplishment, attachment to the business community, and preoccupation with civic political affairs.”); Papke, supra note 17, at 38 (discussing disproportionate effects of the Canons, particularly limitations of expertise claims and advertising, on solo practitioners who serviced working class people). See generally Harry Cohen, Ambivalence Affecting Modern American Law Practice, 18 Ala. L. Rev. 31, 31 (1965) (“Many rules and principles which purport to guide professional conduct today are based on the premise that the American lawyer is in the same economic and professional environment as his predecessors who practiced in the nineteenth century or as barristers in the English system.”); Donald T. Weckstein, A Re-Evaluation of the Canons of Professional Ethics—Evaluated, 53 Tenn. L. Rev. 176, 180 (1966).

32. See Carle, supra note 16, at 16 (noting that all fourteen final committee members were members of the “social and economic elite of the profession” and were exclusively white, Anglo-Saxon, and Protestant).

33. See Gillers & Simon, supra note 19, at 617-18; Hazard et al., supra note 18, at 14.
existing Canons. However, attempts to reword the Canons "became an extended search for the full meaning of professional responsibility in the context of modern day society, a search that culminated in the formulation of the [Model] Code." As such, the Committee began a multi-year journey of deliberations that eventually resulted in the Model Code of Professional Responsibility.

B. From Canons to Code: Operationalizing Theory

Wright Committee members were mindful of the political climate and contemporary changes facing the legal profession. The preface to the Model Code makes explicit, the Model Code was drafted in direct response to contemporary practice. It noted that the "changed and changing conditions in our legal system and urbanized society require new statements of professional principles." New demographics of people were joining or about to join the profession in force. The "[r]ecruitment into the profession was affected by programs reaching out to racial minorities and women, whose assimilation into law practice became both a norm of public policy and a legal duty." The profession was also growing quickly in size, partially in response to the expansion of the administrative state. The Committee was aware of the urgency behind the need to adapt. It emphasized this in the Code:

34. See Wright, supra note 10, at 5 ("A completely changed document was not envisioned.").
35. Id. at 6.
36. The chair of the Committee, Edward L. Wright, stated that the Model Code was a "substantial improvement" over the Canons precisely because "it is the result of a thorough review of the functions of lawyers in modern-day society." Id. at 17.
37. MODEL CODE OF PROF'L RESPONSIBILITY Pmbl. n.5 (1980) ("The law and its institutions change as social conditions change. They must change if they are to preserve, much less advance, the political and social values from which they derive their purpose and their life. This is true of the most important of legal institutions, the profession of law. The profession, too, must change when conditions change in order to preserve and advance the social values that are its reasons for being, ") (quoting Elliott E. Cheatham, Availability of Legal Services: The Responsibility of the Individual Lawyer and the Organized Bar, 12 U.C.L.A. L. REV. 438, 440 (1965)).
38. Id. at Preface.
40. Hazard, supra note 7, at 1259.
41. See BARBARA A. CURRAN, WOMEN IN THE LAW: A LOOK AT THE NUMBERS 8-9 (1995) (noting that because of a steep increase in total numbers of law students, increased percentages of women law students did not displace male students in the total number of law jobs available to them); Hazard, supra note 7, at 1259 n.109.
The advances in natural science and technology are so startling and the velocity of change in business and in social life is so great that the law along with the other social sciences, and even human life itself, is in grave danger of being extinguished by new gods of its own invention if it does not awake from its lethargy.

The extent to which societal shifts were on the minds of the drafters of the Code is evident in the Model Code’s multiple references to evaluating lawyer’s roles in response to historical context and fluid circumstances. The preamble of the Model Code also reminded lawyers that “a lawyer must with courage and foresight be able and ready to shape the body of the law to the ever-changing relationships of society.” Acknowledging the momentous changes in society, the Code identifies the difficulties this places on the bar itself:

Changing times produce changes in our laws and legal procedures.

We have undergone enormous changes in the last fifty years within the lives of most of the adults living today who may be seeking advice. Most of these changes have been accompanied by changes and developments in the law. Every practicing lawyer encounters these problems and is often perplexed with his own inability to keep up, not only with changes in the law, but also with changes in the lives of his clients and their legal problems.

Thus, mindful of the need to transform and the social pressures at play in an “urbanized society,” the Wright Committee embarked on drafting a Model Code that met these needs, yet was “designed to be acceptable to the profession” as it currently stood. The Wright Committee itself was hardly an anti-establishment group. Although the Committee chair, Edward L. Wright, noted that the twelve members of the Committee represented “a broad spectrum of the profes-

42. Model Code of Prof’l Responsibility EC 8-1 n.1 (1980).
43. See, e.g., id. at EC 2-7 (“Changed conditions, however, have seriously restricted the effectiveness of the traditional [attorney] selection process.”); id. at EC 2-10 (“Because technological change is a recurrent feature of communications forms, and because perceptions of what is relevant in lawyer selection may change, lawyer advertising regulations should not be cast in rigid, unchangeable terms.”); id. at EC 8-1 (“Changes in human affairs and imperfections in human institutions make necessary constant efforts to maintain and improve our legal system.”).
44. See id. at Pmb.
45. Id. at EC 6-1 n.1 (internal quotations omitted).
46. Id. at Preface.
47. Hazard, supra note 7, at 1252.
sion,” others might categorize it as a fairly narrow slice of society reflecting the traditional demographics of the profession.49 The Committee included the Associate Dean of Harvard Law School,50 the Dean of Northwestern Law School,51 one former Supreme Court Justice,52 and nine current practitioners in private practice (including two former ABA presidents).53 Wright himself was a lawyer from private practice in Arkansas where he was a longtime active member of the ABA, a leader of the American College of Trial Attorneys, and a prospective future president of the ABA.54 He was known for being “outwardly conventional—indeed formal—in dress, conversation, and deportment.”55 The Committee did not include women or members of minority racial or ethnic groups. The Wright Committee’s membership is notable as the process of drafting the rules was private; the Committee deliberated in closed meetings, no interim drafts were published or circulated, and no hearings were held.56 During the final years of the process, a young female attorney acted as an assistant reporter to the Committee, but there is no indication that she was consulted for substantive input.57

The Wright Committee faced unique challenges. Unlike the drafters of the Canons in 1908 who were drafting rules for a comparatively homogenous membership, the Wright Committee was crafting rules for an increasingly diverse group of lawyers. The drafters likely did not share many personal connections or common experiences with these new entrants.58 Recognizing that there would be multiple views of what is moral conduct in the emerging bar, “the issue quickly becomes, whose conscience and whose ethical standards are to con-

49. Id. at 2.
50. Id. at 2 n.3.
51. Id.
52. Id. at 2 n.1.
54. See Wright, supra note 10, at 2.
55. Hazard, supra note 7, at 1252.
56. See id. at 1253.
57. John F. Sutton, a University of Texas law professor and former chairman of the Texas State Ethics Committee, acted as the official reporter. For the last two years of the process, Sutton was assisted by a recent University of Texas School of Law graduate, a young woman who came to take a very active role in the civil right struggle as the lead counsel in *Roe v. Wade*, Ms. Sarah Weddington. See Wright, supra note 10, at 2-3. Weddington also actively aided chairman Wright in publications related to the Code. Id.
58. As such, the disciplinary rules “functioned as a statute defining the legal contours of a vocation whose practitioners were connected primarily by having been licensed to practice law.” Hazard, supra note 7, at 1251.
trol?‖ 59 Unable or unwilling to engage this question, the Wright Committee devised an alternative to direct confrontation: it created a system to ignore such differences and attempt to insulate the workplace from value pluralism. Specifically, the drafters of the Model Code (1) separated ethical considerations from a “floor” of acceptable conduct (known as disciplinary rules), thereby removing any mandatory discussion of morality (as opposed to legality) from general discourse or debate on the bar level, and (2) strengthened and operationalized a commitment to neutral partisanship that facilitated the removal from the workplace of lawyers’ identities as people with moral viewpoints. 60

The Committee’s work culminated in a draft of the Model Code that was approved by the ABA House of Delegates in the summer of 1969 and went into effect in 1970. 61 The Model Code was a significant structural departure from the Canons: it was regulatory in nature and bifurcated for the first time rules from ethical considerations. 62 Instead of a prosaic format, the Wright Committee structured the Model Code in statutory-style tiers. The Code grouped rules according to nine “Canons” that “are basic to the proper functioning of the legal profession in modern society.” 63 Each Canon had a one-sentence-long overarching statement that began with the discretionary qualifier “A lawyer should” and then proceeded to a statement such as “assist in preventing the unauthorized practice of law” or “assist in maintaining the integrity and competence of the legal profession.” 64 Following each of the header-style Canons were binding mandatory “Disciplinary Rules” (DRs)—created, as their name suggests, to discipline 65—and accompanying, non-binding “aspirational” “Ethical Considerations” (ECs) that were provided for guidance. 66

59. Sutton, supra note 6, at 260 (official reporter for the Code writing on reasons supporting the Code’s development).

60. In drafting the Model Code, the Committee deliberately sought “a complete separation . . . between the inspirational and the proscriptive,” which they viewed as a “substantial improvement.” Wright, supra note 10, 17; see also Sutton, supra note 6, at 260 (“The creation of the ethical considerations-disciplinary rules bifurcation ends the structural difficulty. Now it can be stated that the law providing specific, authoritative standards for the advocate includes the disciplinary rules.”).

61. See Gillers & Simon, supra note 19, at 617.

62. See Lisa G. Lerman & Philip G. Schrag, Ethical Problems in the Practice of Law 46 (3d ed. 2012); Hazard, supra note 7, at 1249-60; Sutton, supra note 6, at 258.

63. See generally Model Code of Prof’l Responsibility Table of Contents (1980).

64. Id. at Canon 1, 3.

65. See Sutton, supra note 6, at 258.

66. Id. (stating that ethical considerations “are designed to ‘appeal to the reason and understanding of the lawyer’ and to give guidance in those areas in which the lawyer is free to exercise his own conscience without compulsion of law” (quoting John F. Sutton, Jr.,
The Model Code quickly became the default measure of attorney misconduct in federal courts; within a year of being presented to the ABA, seventeen states had adopted the Model Code, and the large majority of remaining states soon fell in line.\(^67\)

Reformers heralded the Model Code as a moderate victory.\(^68\) In retrospect, the Model Code that emerged in 1969 has been viewed as an improvement from the preceding Canons.\(^69\) To many commentators then and now, the Model Code facilitated modern practice and the diversifying demographic composition of the bar by increasing transparency, modifying rules to allow more flexibility regarding referrals and advertising,\(^70\) and discussing a commitment to pro bono work.\(^71\) The overall tone of the Model Code, particularly the preamble, echoed with civic aspirations.\(^72\) The Model Code remained at play until the mid-eighties when its replacement, the ABA Model Rules of Professional Conduct, was adopted by the ABA and eventually most states.\(^73\) The Model Rules, like the Model Code before it, are code-like, legally enforceable, and also set a baseline of lawyer conduct grounded in neutral partisanship.

III. CODIFICATION AND STRUCTURE

The advent of the Model Code was a watershed moment for American legal ethics. With it, the American bar eschewed a duty to engage in broad discussions of the lawyer’s moral role in civil society for a limited inquiry into what the proper regulatory rules should be to

---

\(^67\) See LERMAN & SCHRAF, supra note 62, at 46; Wright, supra note 10, at 1.


\(^69\) See LERMAN & SCHRAF, supra note 62, at 46-47 (noting that “the Model Code, however, was an important advance”); Sutton, supra note 6, at 266 (“The Code of Professional Responsibility represents (in my perhaps biased judgment) a giant step forward in the efforts of the legal profession to improve its ethical climate.”).

\(^70\) See MODEL CODE OF PROF’L RESPONSIBILITY DR 2-101(B), (D), (I) (1980); id. at DR 2-103(B), (D); id. at DR 2-140(A). Note, the greatest changes in this area came in subsequent revisions in response to the Supreme Court’s ruling that lawyer advertising received First Amendment protection as commercial speech in Bates v. State Bar of Arizona, 433 U.S. 350 (1977).

\(^71\) See id. at DR 2-101(B), (D), (I); id. at DR 2-103(B), (D); id. at DR 2-140(A).

\(^72\) See id. at Pmbl.

\(^73\) See GILLERS & SIMON, supra note 19, at 617. The Model Code was superseded by the Model Rules of Professional Conduct (Model Rules), which were adopted by the ABA in August 1983. See House of Delegates Proceedings, 108 ANN. REP. A.B.A. 763, 778 (1983).
monitor day-to-day legal practice. This moment was pivotal because it officially relegated issues of morality and lawyering to the “academic” or “personal” sphere and deemed such discussions irrelevant, inappropriate, or even illegitimate in the context of the practicing bar. While the rules did not bar discretionary consideration of morality, there was no requirement to weigh morality. The moral status quo became the elephant in the room, obscured by the immediate mandatory task of complying with disciplinary rules. Codification is often associated with transparency and its attendant virtues: open access, due process, and competition of ideas. Less emphasized, however, is the impact of codification on operationalizing and enforcing norms that were previously unenforced. Enter the Model Code.

A. Rules Adopted; Ethics Orphaned

The Model Code’s structure renders discussions of ethics and professional conduct conceptually and practically distinct. The Code prioritized conduct rules over obligations to consider ethics. Conduct rules received the distinction of being enforceable and therefore im-

74. See LERMAN & SCHrag, supra note 62, at 46 (“The codification of the law governing lawyers in the 1960s marked a major change in the structure and content of the ethical rules.”); Hazard, supra note 7, at 1251 (“The transformation of the norms of professional conduct [into an enforceable legal code] was principally effected by the ABA’s Code of Professional Responsibility in 1970.”).

75. This is not to say that discourse under the Canons actually centered more on substantive ethics. Certainly we have no evidence of a robust moral dialogue happening at the bar level under the Canons. In fact, there is scholarship indicating that civic-mindedness was actually in short supply. See generally Norman W. Spaulding, The Myth of Civic Republicanism: Interrogating the Ideology of Antebellum Legal Ethics, 71 FORDHAM L. REV. 1397 (2003). The key difference with the Code, however, is that the Canons aspired to a notion of professional behavior that included and expected discussions of morality, regardless of whether those ideals were in fact realized.

76. See Hazard, supra note 7, at 1249 (“[T]he legal profession’s narrative and the core ethical rules, as pronounced in the 1908 Canons, has been preserved . . . . However, the form in which those rules are expressed has changed dramatically. What were fraternal norms issuing from an autonomous professional society have now been transformed into a body of judicially enforced regulations.”).

77. Indeed, according to the Committee Chair Wright, the shift to the Code format in the Model Code was partially animated by a desire to create a fair system to incorporate new members into the established norms of the legal profession: “The Rules are drafted in the form of statutes with specificity and clarity sufficient to meet due process requirements of disciplinary actions.” Wright, supra note 10, at 2.

78. See Hazard, supra note 7, at 1250-51 (noting that integrating the bar as an aegis of the courts allowed for “intensified disciplinary enforcement, including the sanctions of disbarment and suspension”).

79. This distinction and the inclusion of ethical consideration with rules may have been inspired by the Joint Conference Report, which noted that “[u]nder the conditions of modern practice it is particularly necessary that the lawyer should understand, not merely the established standards of professional conduct, but the reasons underlying these standards.” Fuller & Randall, supra note 26, at 1159.
portant. By structurally bifurcating ethical considerations from the disciplinary rules, the Model Code made clear that ethical issues (to the extent that they were discussed at all) were non-binding and “aspirational.” In adopting this format, the Model Code set the stage for minimizing discussions of morality among practicing lawyers and law students. Post-Model Code, the lawyerly inquiry is not “What is the right thing or the professional thing to do?” but “Do the Disciplinary Rules sanction doing or not doing A, B, or C?” Morality is relegated to an issue of private contemplation, rather than a topic for group analysis. As such, the Model Code denied new entrants to legal practice the opportunity to discuss and change the moral norms governing lawyering as a whole. Instead, professional discourse focused on the floor provided by regulatory rules, rather than the moral ceiling.

In doing so, the bar was able to avoid a plurality of views concerning a lawyer’s core ethical duties (thus side-stepping the Wright Committee reporter’s question, “whose conscience and whose ethical standards are to control?”). Instead, it froze the conversation as it stood. Bifurcating moral issues from professional conduct allowed current members of the bar to avoid uncomfortable conversations with dissimilar colleagues. Meanwhile, discussions of the moral role of lawyers and related duties to society would become increasingly complicated as the bar diversified.

81. See Benjamin H. Barton, The ABA, the Rules, and Professionalism: The Mechanics of Self-Defeat and a Call for a Return to the Ethical, Moral, and Practical Approach of the Canons, 83 N.C. L. REV 411, 411 (2005) (arguing that the goals of the Model Code and Model Rules’ ‘minimalist’ project and the “broadly ethical” project conflict and therefore “failed largely because the profession has divided what was once the single unifying goal for bar associations and lawyer regulators—providing moral, ethical, and practical guidance on how to practice law—into two quite distinct, and in some ways contradictory, goals, thus undercutting the entire project”); David Luban & Michael Millemann, Good Judgment: Ethics Teaching in Dark Times, 9 GEO. J. LEGAL ETHICS 31, 41-46 (1995) (arguing that the regulatory focus of the codes has removed morality from ethics).
82. See Hazard, supra note 7, at 1241.
83. Some would argue that moral discussions still happen but in relation to the content of the rules, if not generally. I agree that (as with any law or code) moral arguments may be raised in questioning the legitimacy of rules governing lawyer conduct. However, these critiques center on the moral value of the specific rule at issue, and there is no requirement that morality be discussed generally amongst the bar or with clients or that morality be reflected in the rules or conduct. Law often has a disconnect with moral judgment. But we do not call law ethics. Moreover, discussions regarding the reform of rules will happen predominately in bar committees and subcommittees and in some cases among regulators (Sarbanes-Oxley and related reforms), the elite, or academically inclined, not amongst lawyers generally.
84. Sutton, supra note 6, at 259-60 (official reporter for the Code writing on reasons supporting the Code’s development).
For example, a key point of contention facing the legal community in the 1960s was the moral legitimacy of civil disobedience. During the same years when the Wright Committee met and drafted the Code, both ABA President Powell and former Justice Whittaker (a member of the Wright Committee) spoke publically against civil disobedience as a means of social change.\(^8^5\) Instead, they called for stronger government enforcement of criminal laws. One of their concerns was that a rise in disrespect for the law and general lawlessness would undermine “the good order and morality of our society.”\(^8^6\) Whittaker in particular was skeptical of demonstrations and questioned the motivations of civil rights protesters. He argued that “certain self-appointed racial leaders, doubtless recalling the appeasements and, hence, successes of that earlier conduct, have simply adopted and used those techniques in fomenting and waging their lawless campaigns which they have called ‘demonstrations.’”\(^8^7\)

Justice Whittaker stated that “we must always strive to eliminate injustice and discrimination.”\(^8^8\) However, he disagreed with non-violent civil disobedience demonstrators about how to achieve that goal. He argued that “we must do so by orderly processes in the legislatures and the courts, and not by defying their processes and actions, nor by taking the laws into our own hands.”\(^8^9\) Tellingly, Justice Whittaker’s writings on the subject lack awareness of institutional bias. He urged minorities to make use of the court system without acknowledging that the laws to which minorities are expected to show allegiance were often formed without their input and in violation of fundamental constitutional and moral principles.\(^9^0\) Likewise, there is no apparent cognizance of how judicial entities themselves are entrenched in their own norms, norms that may foreclose the possibility of meaningful redress through conventional legal processes. Rather,

\(^{85}\) Concerned about civil disobedience’s ability to “seriously threaten the breakdown of law, order, and morality,” Charles Whittaker, in an address originally given to the Tennessee Bar Association, called for “impartial, evenhanded, vigorous, swift and certain enforcement of our criminal laws, and the real and substantial punishment thereunder of all conduct that violates those laws.” 111 CONG. REC. 15101-03 (1965) (introduced into the record by Sen. Strom Thurmond); id. at 15103 (arguing that “[n]o ‘end’ . . . however worthy [can ever] justify resort to unlawful means” and that “America needs a genuine revival of respect for law and orderly process . . . a new impatience with those who violate and circumvent the laws, and a determined insistence that laws be enforced” (alterations in original)).

\(^{86}\) Id. at 15102.

\(^{87}\) Id.

\(^{88}\) Id. at 15101-03.

\(^{89}\) Id. at 15103.

\(^{90}\) See id.
Justice Whittaker argued that complying with the law and upholding legal institutions is the sole legitimate path to effectuate change.91

In the above example, the moral inquiry asks whether a lawyer owes a moral duty to follow the law even when the law is unjust. Members of the Committee viewed this as a potential clash, one where some lawyers would be more willing to assert that unjust laws are not entitled to automatic compliance. Such lawyers could argue that compliance with an unjust law is, in fact, immoral. Under the Model Code, practicing lawyers need not have this discussion or resolve this issue with their colleagues. The focus on the disciplinary rules made clear that lawyers need only discuss what is proper conduct in the service of clients, not whether or not clients ought to be prioritized over other societal obligations or whether lawyers should in fact consider the morality of the laws they are instrumental in implementing.

The avoidance of moral discussions in the professional context has continued to this day. The discussion of legal ethics in law schools is almost entirely rule-based.92 “Legal Ethics” as a course name is often a misnomer for continuing legal education and law school courses that teach not frameworks for considering the moral implications or obligations of practice, but, rather, teach the predominantly code-based law of lawyering.93 To the extent that lawyers consider the bounds of their duties to society, their clients, and the legal system, it is primarily, and perhaps only, in terms of complying with formal laws and regulations.94 From the ratification of the Model Code on-

91. See id. (“We must . . . seek redress in the courts rather than in the streets.”).

92. This is partially due to the sweeping mandate of the ABA’s accreditation provisions for the mandatory ethics course in law school which requires treatment of “the history, goals, structures, values, and responsibilities of the legal profession and its members.” A.B.A., THE 2012-2013 ABA STANDARDS AND RULES OF PROCEDURE FOR APPROVAL OF LAW SCHOOLS Standard 302(a)(5), at 15 (2012).

93. In required CLE and law school courses on “legal ethics” there is little need, and certainly no requirement, to discuss morality. Of a survey of thirty-one textbooks on the subject of legal ethics and professional responsibility on the current market in 2013, fewer than half even use the term “ethics” in their titles. See also Maksymilian Del Mar, Beyond Text in Legal Education: Art, Ethics, and the Carnegie Report, 56 Loy. L. Rev. 955, 976-77 (2010) (noting that professional responsibility and ethics courses in law school do not address moral concerns but focus instead on the regulation of law practice); Deborah L. Rhode, Ethical Perspectives on Legal Practice, 37 Stan. L. Rev. 589, 649 (1985) (“Legal ethics should be taught as ethics, not as etiquette or statutory exegesis. Law school courses and bar examinations that demand rote memorization of official standards merely trivialize the subject matter.”).

wards, the term “ethics” is not even in the titles of the ABA rules governing the legal profession.  

The division of ethics from rules in the Model Code presented morality as an issue outside of the lawyer’s professional responsibility or training, a personal discretionary issue, without a clear place in professional discourse or the workplace. While the Canons’ moral validity was highly flawed, the ideal of lawyering perpetuated by the Canons required a consideration of moral issues. An attorney could not be a good lawyer and fail to consider (and act on) morality. In contrast, the neutral-partisan model in the Model Code and in the Model Rules today allows an individual to be a good lawyer and not weigh or act on morality at all. The model of lawyering that comes from the Model Code sets the expectation that lawyers will act as “amoral” agents. Yes, lawyers can exercise discretion to go above and beyond what is required of them as lawyers and consider morality (as many do). But the fact is that they are not required to. This shift relegated legal ethics qua ethics to the ivory tower of academia and personal, rather than professional, discourse.

Taking moral discussions off the table for lawyers also meant that lawyers unsympathetic to new viewpoints avoided the need to reconsider or modify their own practice. The proper role of lawyers in society was beyond group revision. Moral condemnation of the basic role of lawyer as client conduit was out of place. Instead, the focus on day-to-day conduct allowed the existing bar to shield the client-centric neutral-partisan model from serious attack while focusing discussion on disciplinary rules. This move conceded broad conceptual and moral issues to the status quo. The neutral-partisanship model allowed members of the bar to avoid awkward discussions regarding their personal views with dissimilar colleagues because it assumes that a lawyer’s personal morality remains separate from achieving client goals.


96. To be clear, this is not a call to reinstate the Canons. Others have called more generally for return to the Canons, whereas I am making a more limited claim. See Barton, supra note 81, at 434-36.

97. See infra notes 118-21 and accompanying text.

98. As I discuss in detail infra Part III, there is no such thing as an amoral lawyer; the amoral lawyering ideal is imbued with rich moral judgments regarding agency, duty, societal norms, and expectations regarding human behavior.

99. It is worth noting that new members to the bar from outside of traditional professional circles are not spared discomfort under this model; rather, their discomfort may be increased in a system where topics available for discussion or redress forgo challenges to moral concerns in the context of professional practice.
B. Enforceability: The Model Code and Discipline

As the demographic composition of the legal profession grew more inclusive, longstanding members of the bar could no longer rely on common experiences of education and upbringing, unspoken norms, or societal connections to predict (and control) the behaviors of other lawyers who moved in different social circles.\(^{100}\) The Model Code created a national code of conduct with mandatory regulations specifically designed to increase enforceability over parties outside the ambit of the social circles of the existing bar.\(^{101}\) In doing so, the Code “transformed the dominant bar associations directly involved in bar discipline from private clubs into quasi-governmental organs.”\(^{102}\) As such, the drafting of the Model Code did more than fend off encroachment by the court on traditional self-regulation;\(^{103}\) it revealed the existing bar’s increasing skepticism of the rapidly growing lawyer population’s ability to conform to norms without external consequences.

Newly anointed standards were given teeth. In the pre-Code era, “[t]he threat of professional discipline was virtually non-existent, as long as the lawyer did not commit a felony or a similarly egregious offense. The threat of legal malpractice recovery or of a remedy such as disqualification for a conflict of interest was almost as equally remote.”\(^{104}\) The disciplinary norms changed drastically post-Model Code. The Model Code provided a disciplinary workhorse to keep lawyers in line with dominant lawyering norms, rather than facilitating revision of these norms. Since 1969, lawyer regulation through the bar, courts, and malpractice litigation has risen significantly.\(^{105}\) While some of this increase can be attributed to doctrinal shifts in

\(^{100}\) The Model Code’s disciplinary rules “functioned as a statute defining legal contours of a vocation whose practitioners were connected primarily by having been licensed to practice law.” Hazard, supra note 7, at 1241, 1244, 1251, 1260 (going on to note that “[b]y the 1980’s, the bar had become a ‘community’ of strangers”).

\(^{101}\) See Zacharias, supra note 7, at 1 (“It was not until 1969, with the advent of the Model Code of Professional Responsibility, that American jurisdictions began to take the function of regulating lawyers seriously.”).

\(^{102}\) Wolfram, supra note 2, at 217; see also Hazard, supra note 7, at 1241 (“[T]he legal profession’s narrative and the core ethical rules, as pronounced in the 1908 Canons, has been preserved and largely unchanged. . . . However, the form in which those rules are expressed has changed dramatically. What were fraternal norms issuing from an autonomous professional society have now been transformed into a body of judicially enforced regulations.”).

\(^{103}\) In the 1960s there was increased pressure from courts on the self-governance norm as they increasingly took an increased interest in shaping the profession’s governing rules. Hazard, supra note 7, at 1242.

\(^{104}\) Wolfram, supra note 2, at 207.

\(^{105}\) “While the absence of meaningful records precludes the generation of statistics of the extent of lawyer discipline prior to 1970, my distinct impression, in agreement with the bar’s self-assessment, is that there was much less regulation compared to today.” Id. at 206.
third party standing, the growth of a plaintiffs’ legal malpractice bar may also indicate a sense that the presence of a law-like code clarified the breach standards needed for conformity with the customs of the profession. More than forty years later, lawyers now comprise a highly regulated profession subject to direct statutory regulation and self-regulation as well as various common law regulations.

IV. STANDARDIZING THE STANDARD CONCEPTION

The previous sections established a history of ideas—why neutral partisanship was adopted as a dominant model for lawyering in modernity (to facilitate pluralism) and what form that adoption took (the bifurcation of rules from ethics). The historic roots of the Model Code indicate a perceived need to avoid conflict and excise debates about morality from the legal profession in order to facilitate demographic plurality at the bar. This Part explores the conceptual and practical shortcomings of that choice by problematizing arguments that neutral partisanship is (1) anti-hierarchical and pro-client and (2) neutral to moral judgments.

The Model Code does not (and cannot) remove moral judgment from the code. Rather, by adopting and strengthening a commitment to neutral partisanship, the Model Code preserved the moral choices that predated the Code and limited the ability to change or alter those choices moving forward. Despite its appeal to universality, neutrality is not neutral. It favors one approach to lawyering over all others. Because it is viewpoint specific but does not facilitate dialogue on the content of that viewpoint, neutral partisanship deemphasizes and delegitimizes the individual moral contributions of new diverse lawyers by limiting value pluralism.

As such, the desire to avoid conflict in the bar did not eliminate moral conflict. Instead, the Model Code pushes conflict out of sight, where it is harder to investigate. Such suppressed struggles cause


107. See John Leubsdorf, Legal Malpractice and Professional Responsibility, 48 RUTGERS L. REV. 101, 102 (1995) (“Increasingly, professional ideals have been turned into enforceable law, and self-regulation by the organized bar has become regulation by courts and legislatures.”); Russell G. Pearce, The Professionalism Paradigm Shift: Why Discarding Professional Ideology Will Improve the Conduct and Reputation of the Bar, 70 N.Y.U. L. REV. 1229, 1275 (1995) (noting that the existence of disciplinary regulation and substantial regulatory law, including “tort law, criminal law, agency law, and securities law,” belies the notion that lawyers are self-regulating); Wolfram, supra note 2, at 206; Zacharias, supra note 7, at 1 (“The publication of the Model Code of Professional Responsibility was a watershed event beginning a flood of ethics regulation that has yet to subside.”).

significant harm to lawyers as individuals as well as to relationships between lawyers and between lawyers and clients. Neutral partisanship is also ill-suited to help a diverse bar that services diverse clients to negotiate real moral and practical dilemmas. This Part concludes that the Model Code, as the basis of our modern legal ethics system, does not adequately respond to demographic and value pluralism. As such, it cannot be justified on the terms of its adoption.

A. Neutral Partisanship: A Primer

Neutral partisanship is typically called the “standard conception” of the lawyer’s role, amoral lawyering or, in the pejorative, the lawyer as a “hired gun.” Partisanship is the concept that the lawyer, within the boundaries of the law, “is committed to the aggressive and single-minded pursuit of the client’s objectives.” Partisanship exists primarily in adversarial systems. Neutrality is a distinct concept; it requires that lawyers refrain from moral judgment over the lawful ends of the client or the lawful means used to attain those goals.

Under the standard conception, lawyers are “role-differentiated.” This means that lawyers must confine their individual moral views to their personal roles in life (outside of the workplace). Ideally, law-


111. This Article does not interrogate the adversarial or partisan nature of the standard conception’s “neutral partisan” norm, but instead focuses on the neutrality aspect of this concept. Many scholars have critiqued the pitfalls of partisanship and particularly the use of the adversarial system. See, e.g., DAVID LUBAN, LEGAL ETHICS AND HUMAN DIGNITY 19-64 (2007) (revising Luban’s 1983 essay and critiquing the premise that the adversarial system better pursues truth or the client’s interest); Carrie Menkel-Meadow, The Trouble with the Adversary System in a Postmodern, Multicultural World, 38 WM. & MARY L. REV. 5 (1996). But see MONROE H. FREEDMAN, LAWYERS’ ETHICS IN AN ADVERSARY SYSTEM 9, 12 (1975) (championing the adversary system as protective of individuals’ fundamental rights and emphasizing the importance of partisanship and neutrality); STEPHAN LANDSMAN, THE ADVERSARY SYSTEM: A DESCRIPTION AND DEFENSE (1984) (advocating for the adversary system). However, these concepts are analytically distinct: concepts of partisanship and neutrality need not be married with one another. This Article focuses on problems imbedded with neutrality and role morality, not partisanship.


113. See, e.g., Charles P. Curtis, The Ethics of Advocacy, 4 STAN. L. REV. 3, 15-16 (1951) ("[Lawyers] are not dealing with the morals which govern a man acting for himself, but with the ethics of advocacy. We are talking about the special moral code which governs a
yers leave their professional role devoid of moral judgment and remain preoccupied only with questions of legality.\textsuperscript{114} As such, neutral partisanship provides moral non-accountability for lawyers: the client, not the lawyer, is culpable for the moral worth of ends sought in a representation.\textsuperscript{115} Many refer to this concept as “role morality” since morality, or at least moral accountability, is cabined to specific roles.\textsuperscript{116} Neutral partisanship in some form is a time-honored concept, one that some scholars trace back as far as the early 1800s.\textsuperscript{117}

The Canons’ version of neutral partisanship was qualified and favored lawyers employing their own morality in the context of client advocacy. While the Canons required lawyers to act as zealous advocates for their clients, they also cautioned that a lawyer “must obey his own conscience and not that of the client.”\textsuperscript{118} The Canons reminded lawyers that “no client has a right to demand that his counsel shall be illiberal, or that he do anything therein repugnant to his own sense of honor and propriety.”\textsuperscript{119} Moreover, under the Canons, a lawyer had a duty to “impress upon the client and his undertaking exact compliance with the strictest principles of moral law.”\textsuperscript{120} The Canons concluded that a lawyer finds their “highest honor in a deserved reputation for fidelity to private trust and to public duty, as an honest man and as a patriotic citizen,” not exclusively in the service of a client.\textsuperscript{121}

The Model Code, in contrast, made a much stronger commitment to neutral partisanship.\textsuperscript{122} In it, the Wright Committee sought to eliminate what it perceived as a key “dilemma” imbedded in the Canons who is acting for another. Lawyers in their practice—how they behave elsewhere does not concern us—put off more and more of our common morals the farther they go in a profession which treats right and wrong, vice and virtue, on such equal terms.

\begin{itemize}
\item \textsuperscript{114.} See David Luban, \textit{Misplaced Fidelity}, 90 TEX. L. REV. 673, 673-74 (2012) (describing the position of W. Bradley Wendel in \textit{Lawyers and Fidelity to Law} regarding the proper role of a lawyer).
\item \textsuperscript{115.} See Pepper, \textit{supra} note 11, at 614 (“C]onduct by the lawyer in service to the client is judged by a different moral standard than the same conduct by a layperson. . . . [I]t is the client who is morally accountable, not the lawyer.”).
\item \textsuperscript{116.} Luban, \textit{supra} note 114, at 674 (noting that the central question regarding role morality is, “how can it be that her professional role might require a lawyer to do things that would be morally forbidden to a non-lawyer?”).
\item \textsuperscript{117.} DARE, \textit{supra} note 110, at 6 (2009) (discussing Lord Brougham’s famous speech in defense of Queen Caroline, pronouncing that in the discharge of his duty a lawyer “knows but one person in all the world, and that person is his client”).
\item \textsuperscript{118.} \textit{Final Report}, \textit{supra} note 18, at 579.
\item \textsuperscript{119.} \textit{Id.} at 582.
\item \textsuperscript{120.} \textit{Id.} at 584.
\item \textsuperscript{121.} \textit{Id.} at 584.
\item \textsuperscript{122.} See Papke, \textit{supra} note 17, at 43 (“Framers of the code reasserted the responsibility of individual lawyers to individual clients. Established ethical presumptions died hard.”).
\end{itemize}
ons: "whether to represent [a lawyer's] client in accordance with [the] law or in accordance with the lawyer's own moral viewpoint."123 While the official reporter of the Model Code, John Sutton, admitted that there was "considerable professed support" for the view that a lawyer should "obey his own conscience,"124 he dismissed it as a "standardless maelstrom" governed by "whims, prejudices, emotional caprices, and predilections."125

The Model Code avoided this concern by making only very limited mention of moral responsibility on the part of lawyers. These mentions appear in the non-binding "ethical considerations," not the mandatory disciplinary rules, where the ethical considerations merely state that it is "desirable" for a lawyer to point out morally just outcomes to clients.126 The clients themselves decide the course of action. This side-steps a lawyer's moral accountability.127 For example, some scholars have argued that the Model Code forbids trial lawyers to even express an opinion as to the moral value of their client's case.128 Even in the area of deciding whether or not to take on certain clients, where lawyers under the Canons had previously enjoyed an unqualified right to refuse clients,129 the Model Code admonished such behavior. It declared that "a lawyer should not lightly decline proffered employment."130

Additional commitments to client-centric neutral partisanship are clear in several sections of Model Code. One of the mandatory disciplinary rules states that "[a] lawyer shall not intentionally fail to seek the lawful objectives of his client through reasonably available means permitted by law and the Disciplinary Rules."131 Canon 7 of the Model Code, although not binding in the way the Disciplinary Rules are, reiterates this approach: "A lawyer should represent a client zealously within the boundaries of the law."132 In the non-binding "ethical" authorities, lawyers are advised that "[t]he professional judgment

123. Sutton, supra note 6, at 260.
125. Sutton, supra note 6, at 259.
126. MODEL CODE OF PROF'L RESPONSIBILITY EC 7-8 (1980) ("A lawyer should bring to bear upon this decision-making process the fullness of his experience as well as his objective viewpoint. In assisting his client to reach a proper decision, it is often desirable for a lawyer to point out those factors which may lead to a decision that is morally just as well as legally permissible.").
127. See id. at EC 7-8, 7-9.
129. Final Report, supra note 18, at 583-84 (which emphasizes a lawyer's "right to decline employment").
130. MODEL CODE OF PROF'L RESPONSIBILITY EC 2-26 (1980).
131. Id. at DR 7-101(A)(1).
132. Id. at EC 7-1.
of a lawyer should be exercised . . . solely for the benefit of [the lawyer’s] client and free of compromising influences and loyalties.”\textsuperscript{133}

B. A Modern Moral Justification of Neutral Partisanship

Traditionally, champions of neutral partisanship defend it in pragmatic terms by arguing that neutral partisanship upholds the integrity of the adversarial system.\textsuperscript{134} Under these types of arguments, a lawyer must take a neutral role as an advocate for her client in order to ascertain the truth and safeguard the public interest in the institutional integrity of the overall legal system.\textsuperscript{135} This view argues that allowing lawyers to exercise moral autonomy and judge the goals of the client’s lawsuit undermines the functioning of the adversarial judicial system as a whole, either by usurping the roles of the judge and jury or by undermining the system’s ability to fact find.\textsuperscript{136} In its briefest form, such proponents argue that an adversary system founded on neutral partisanship is the best way to ensure that truth is found and that cases are judged on their merits.\textsuperscript{137} A more cynical argument holds that neutral partisanship is simply a quid pro quo bargain for retaining a monopoly over professional legal services.\textsuperscript{138}

All of these justifications and critiques of neutral partisanship are premised on the basic assumption that existing institutions of law, justice, and government are legitimate themselves and worth protecting. However, in the late sixties and seventies, when the Model Code was written and ratified and subsequent iterations of national rule-making were underway, skepticism of such institutions was wide-

\textsuperscript{133}. \textit{Id.} at EC 5-1.

\textsuperscript{134}. \textit{See} \textit{Freedman, supra} note 111, at 9, 12 (championing the client-centric adversary system as protecting the fundamental rights of individuals and emphasizing the importance of partisanship and neutrality); \textit{Landsman, supra} note 111 (advocating for the adversary system). \textit{See generally} Polk Cnty. v. Dodson, 454 U.S. 312, 318 (1981) (“[Our] system assumes that adversarial testing will ultimately advance the public interest in truth and fairness.”); Mackey v. Montrym, 443 U.S. 1, 13 (1979) (“[O]ur legal tradition regards the adversary process as the best means of ascertaining truth and minimizing the risk of error.”).

\textsuperscript{135}. \textit{See Papke, supra} note 17, at 38 (“Sharswood argued that the lawyer was not responsible for the social utility of the cause he represented. If the lawyer began judging cases on his own, he would be usurping the powers of judge and jury who, more so than lawyers, carried a responsibility to the public at large.”).

\textsuperscript{136}. \textit{See Richard Wasserstrom, Lawyers as Professionals: Some Moral Issues, 5 HUM. RTS. 1, 10-11 (1975) (contending that “[i]f lawyers were to substitute their own private views of what ought to be legally permissible and impermissible for those of the legislature this would constitute a surreptitious and undesirable shift from a democracy to an oligarchy of lawyers”).}

\textsuperscript{137}. \textit{See Fuller & Randall, supra} note 26, at 1160-61.

\textsuperscript{138}. \textit{See Pepper, supra} note 11, at 616.
spread.\textsuperscript{139} Justifications predicated on protecting existing institutions and on the moral integrity of such instructions were questionable—perhaps even untenable. Instead, another moral justification of neutral partisanship was necessary—one resting on contemporary political ideals such as equality, diversity, and individual access to power. It is on these bases that neutral partisanship needed to stake its claim in order to solidify its position as the default model of lawyering in modern America.

Responding to the political and social climate of post-sixties America, modern moral justifications for neutral partisanship emerged grounded in the language and causes of the civil rights movement: individual rights, diversity, and equality.\textsuperscript{140} This moral footing for neutral partisanship is best articulated in Stephen Pepper’s seminal article \textit{The Lawyer’s Amoral Ethical Role: A Defense, a Problem, and Some Possibilities}.\textsuperscript{141}

In the amoral lawyering account, neutral partisanship forwards the cause of the disenfranchised better than a morally active lawyer role because it allows the voice of the client to achieve its full legal ends unimpeded by the lawyer’s own moral viewpoint.\textsuperscript{142} Since these values had inherent moral worth, so too did a role of lawyering conceived to protect such interests.\textsuperscript{143} A lack of sophistication on the part of many clients justifies the lawyer’s amoral role:

\begin{itemize}
\item \textsuperscript{139} See Andrew Kohut, \textit{Forward to PEW RESEARCH CENTER FOR THE PEOPLE & THE PRESS, DECONSTRUCTING DISTRUST: HOW AMERICANS VIEW GOVERNMENT} (1998) (noting how in the sixties and seventies a “healthy skepticism” toward government “deteriorated into an outright distrust”).
\item \textsuperscript{140} See Pepper, \textit{supra} note 11, at 613 (stating that neutral partisanship is morally justified, “primarily upon the values of individual autonomy, equality and diversity”). For the purposes of this Article, I ask the reader to take as given that the ideals of individual autonomy, equality, and diversity are moral goals. While I realize that the moral worth of each of these is subject to debate, this Article considers, if we take Professor Pepper’s categorization of these goals as valid, what does and should follow. See generally GERALD DWORKIN, \textit{THE THEORY AND PRACTICE OF AUTONOMY} 20 (1988) (“[A]utonomy is . . . a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values. By exercising such a capacity, persons define their nature, give meaning and coherence to their lives, and take responsibility for the kind of person they are.”).
\item \textsuperscript{141} Pepper, \textit{supra} note 11, at 613. Professor Pepper’s article is a cornerstone of the modern legal ethics canon. A recent search of LexisNexis revealed that, in addition to being cited in numerous legal ethics anthologies and textbooks, Pepper’s article has been cited over three hundred times since 1986.
\item \textsuperscript{142} Pepper is concerned that lawyers will act as “moral screens” obscuring client autonomy. \textit{Id.} at 621.
\item \textsuperscript{143} Staunch individualism has a long history in American political mythology. The Model Code’s own opening sentence states that “[t]he continued existence of a free and democratic society depends upon recognition of the concept that justice is based upon the rule of law
\end{itemize}
[U]nfiltered access to the law is then available only to those who are legally sophisticated and to those able to educate themselves sufficiently for access to the law, while those less sophisticated—usually those less educated—are left with no access or with access that subjects their use of the law to the moral judgments and veto of the lawyer.\footnote{144}

Under this view neutral partisanship is necessary in a pluralistic society to safeguard the masses from the oppressive oligarchy of the lawyer class.\footnote{145} Justifying neutral partisanship in these terms presented reform-minded lawyers with an argument in favor of strict role differentiation rooted in their own values. The idea that different views and parties would coexist best by seeking a neutral baseline for lawyer conduct has a certain intuitive appeal.

Proponents of neutral partisanship, in service of pluralism, equality, and diversity argued that in a pluralistic society moral lawyers not only may, they must differentiate their personal selves from their professional ones in order to allow various client views to filter into the system: “The lawyer is a good person in that he provides access to the law; in providing such access without moral screening he serves the moral values of individual autonomy and equality.”\footnote{146} Since “liberty and autonomy are a moral good,” so too is neutral partisanship, which is “better than constraint.”\footnote{147} Under this theory, if given a choice, diversity and autonomy are so morally valuable that they trump general pursuits of “right” or “good” conduct.\footnote{148} Thus, “[f]or access to law to be filtered unequally through the disparate moral views of each individual’s lawyer does not appear to be justifiable.”\footnote{149}

However intuitive, this assertion fails upon closer scrutiny. Client-centric neutral partisanship cannot be adequately justified in terms of defending client autonomy, equality, or diversity since it ultimately undermines each of these concepts in important and irreconcilable

\footnote{144}{Pepper, supra note 11, at 619.}
\footnote{145}{See id. at 617 (arguing that neutral partisanship facilitates a diversity of clients to play out their individual autonomy in an unimpeded fashion without “substitut[ing] lawyers’ beliefs”).}
\footnote{146}{Pepper, supra note 11, at 634.}
\footnote{147}{Id. at 616.}
\footnote{148}{In Professor Pepper’s view, equality of access to law is also a “significant value” worthy of such extreme protection and one that the neutral-partisan model champions. Id. at 616.}
\footnote{149}{Id. at 618.}
ways. The following sections examine these shortcomings both conceptually and pragmatically.

C. Neutral Partisanship as a Vehicle for Client Autonomy

The idea that exercises of individual autonomy in accordance with law are an unequivocal moral good is a subject of debate. One danger of this assertion is that it “would conflate the morality of the action with the morality of autonomously having chosen it.” Professor David Luban notes this is particularly problematic because “[s]ome things [that are] legally right are not morally right.” In contrast, amoral lawyering “assumes that the morality is already in the law, that in any important sense anything legally right is morally right.” In order for this to be true, neutral partisanship under the amoral lawyering view would carry an extreme libertarian view that any infringement on individual freedom is so morally repugnant as to outweigh the moral downfalls of the content of that act.

Additional arguments regarding the modern amoral lawyering rationale are grounded in practice; even if an individual makes a choice for a given outcome, if the ability to reach that outcome is fundamentally compromised by her relationship with her lawyer, then the client’s exercise of autonomy does not exist in any meaningful way. This section contends that regardless of perfect access to amoral law-

150. See, e.g., W. BRADLEY WENDEL, LAWYERS AND FIDELITY TO LAW 4, 11 (2010) (arguing that a modified version of neutral partisanship that places fidelity to legal entitlements rather than client interests is essential in a value pluralistic society since law represents “a provisional settlement of these controversies, to enable cooperative action in response to some collective need” and that “[t]here is moral value in doing one’s part to support a socially valuable institution”). I reserve discussion of the issue of whether this modified version of neutral partisanship would succeed in terms of defending pluralism and focus here on assessing the more conventional version of neutral partisanship at play in the Model Code and incorporated into subsequent rulemaking.

151. Id. at 33.


153. Id.

154. For the purposes of this Article, I will assume, as most legal ethics codes do, that good lawyering includes loyalty to one’s client as well as communication and candor between the lawyer and her client. As of July 2014, all jurisdictions in the United States include provisions that outline a duty of loyalty, communication, and candor towards clients in some form. It is a matter of live dispute, however, whether such duties, or any duties, should run to clients versus the law, human dignity, or an overall sense of substantive justice. The amoral lawyering stance discusses individual autonomy in terms of clients, and therefore this Article will approach the standard conception from the point of view that is now, in the post-Code era, most widely accepted—a client-centric model.
yers, the neutral-partisan ideal would continue to be suspect under justifications grounded in diversity, equality, and individual autonomy.\textsuperscript{155}

1. Candor, Loyalty, and Effective Client Service

Many would include zeal as an essential element of good lawyering.\textsuperscript{156} Here lies an inherent weakness in neutral partisanship: it undermines loyalty and the trust of clients in their lawyers. This inhibits the ability of lawyers and clients to have a relationship with a meaningful implementation of loyalty, communication, and candor.\textsuperscript{157} By being willing, for pay, to set aside one’s personal principles and views, lawyers undermine their moral standing and forfeit credibility with laypeople.\textsuperscript{158} Then, those same laypeople are put in the uncomfortable position of having to trust a party they know to be, by all typical standards, untrustworthy. Without the candor that comes with trust, a lawyer cannot build an effective case and relationship with a client. Thus, the amoral lawyering model frequently handicaps a lawyer from effective practice and even client service.

The idea that clients can exercise unimpeded autonomy through a neutral-partisan model of lawyering is undercut by the relationship between trust, empathy, and candor.\textsuperscript{159} People come to lawyers in times of great vulnerability, often when something of critical importance in their lives is awry.\textsuperscript{160} Clients want to trust their lawyers.\textsuperscript{161} Yet a lawyer who complies with her duties as delineated by

\begin{itemize}
  \item \textsuperscript{155} I concede that in a non-adversarial system, neutrality might be more defensible. This Article argues within, and to some extent assumes as given, our current system—an adversarial one grounded in client-centric work. For the reasons hinted at below, I have serious concerns about the moral worth of non-adversarial judicial systems in deliberative democracies, though I would agree with other scholars that client focus should potentially yield to other loyalties and is likely better conceptualized as a duty “to protect the legal entitlements of clients, not advance their interests.” WENDEL, supra note 150, at 6.
  \item \textsuperscript{156} See, e.g., Anita Bernstein, The Zeal Shortage, 34 HOFSTRA L. REV. 1165 (2006). But see LUBAN, supra note 111, at 19-64 (revising Luban’s 1983 essay).
  \item \textsuperscript{157} See, e.g., Irma S. Russell, Keeping the Wheels on the Wagon: Observations on Issues of Legal Ethics for Lawyers Representing Business Organizations, 3 WYO. L. REV. 513, 522 (2003) (“[T]he lawyer who bites his tongue rather than voice the unpleasant argument against a client’s course of action fails more than his own conscience; he fails to fulfill the foundational duty of providing candid legal advice.”).
  \item \textsuperscript{158} Perhaps this is why the media valorizes lawyers who break with professional ethics and instead comport with expectations of common morality. See Robert C. Post, On the Popular Image of the Lawyer: Reflections in a Dark Glass, 75 CALIF. L. REV. 379, 386 (1987).
  \item \textsuperscript{159} The Oxford English Dictionary defines empathy as “[t]he . . . power of projecting one’s personality into [and so fully comprehending the] . . . object of contemplation.” OXFORD ENGLISH DICTIONARY (2d ed. 1989). Since lawyers, under a neutral-partisan ideal, seek to treat their personal selves as irrelevant, empathy is out of place.
  \item \textsuperscript{160} See Pepper, supra note 11, at 615.
  \item \textsuperscript{161} See generally COREY S. SHDAIMAH, NEGOTIATING JUSTICE: PROGRESSIVE LAWYERING, LOW-INCOME CLIENTS, AND THE QUEST FOR SOCIAL CHANGE (2009) (citing
the Model Code can greet them with a complete lack of sympathy or empathy—in fact, they are expected to do just that. Lawyers, by separating their personal selves strictly from their professional selves, appear “heartless,” only in the representation to receive fees and therefore only looking out for their own personal material interest.162 A lack of aligned interests may also impact the ability of a lawyer to engage in competent representation. If a lawyer believes in the case of her client, then she is genuinely motivated to be timely, diligent, and creative and to work hard to reach the client’s goals. Alignment of interests may also support competency, as “[t]here are severe limitations on the extent to which a person, particularly a stranger, can understand with any depth the ends of another without actually sharing those ends.”163 But instead of expecting lawyers only to take on cases that comport with their ordinary moral views, lawyers are expected to represent all clients zealously without believing in the outcomes of their cases.

2. Codifying Trust: The Model Code’s Bind

Neutral partisanship creates the situation where the Model Code must artificially attempt to emulate qualities of actual loyalty-based relationships through piecemeal rubrics. The Model Code states that lawyers must be loyal, diligent, and competent with client’s work and represent the client with zeal.164 The Model Code attempts to deal with the disconnect between role differentiation and traditional loyalty by crafting a set of contorted rules that have only become more painfully strained and convoluted over time. Much of the Code attempts to regulate and create mechanically what can only grow out of a genuine alignment of interests: trust and loyalty.165

162. “Greed” is a common term that arises in public descriptions of lawyers. See Paul F. Teich, Are Lawyers Truly Greedy? An Analysis of Relevant Empirical Evidence, 19 TEX. WESLEYAN L. REV. 837, 847 (2013); see also Marianne M. Jennings, Moral Disengagement and Lawyers: Codes, Ethics, Conscience, and Some Great Movies, 37 DUQ. L. REV. 573, 575-76 (“[M]oral disengagement still creates an ethical pressure cooker from which there is no release.”).


164. See MODEL CODE OF PROF’L RESPONSIBILITY DR 4-101 (1980) (client confidences); id. at DR 6-101 (competence and diligence); id. at DR 7-101 (zeal).

165. See generally id. at Canon 6 (laying out detailed instructions regarding conflicts of interest).
However, trust between people is ordinarily predicated on honesty and mutuality. It requires candid disclosure of beliefs and the mutual vulnerability attendant to that disclosure. No amount of mechanical rulemaking can counter the lack of mutual openness in building trust. Therefore, if the legal profession truly believes that client candor is essential to representation in an adversarial system, then a failure to discuss, disclose, and act in keeping with a lawyer’s autonomy undermines a key aspect of representation. Clients will trust lawyers who are telling them the truth about themselves and being forthcoming with their views. In contrast, the neutral-partisanship model asks clients to do what lawyers themselves are unwilling to do—be open.

Ultimately, rules can only give assurance to clients who have faith in rules, rule of law, and the fair execution of law. Proxy loyalty is little more than a house of cards built upon certain assumptions regarding human experience: that clients will enter a relationship with a lawyer believing that law and rules are reliable, fair, and justly enforced. For many clients, particularly less enfranchised individuals or non-institutional clients, the inverse view of the reliability of rules and law is just as likely as such faith.

Ultimately, client autonomy is best served by lawyers who genuinely believe in clients and their causes. Neutral partisanship divorces lawyers from this rich sense of loyalty that flows organically from aligned interests. Instead, clients must trust in rule of law, the code of lawyers, and the court to enforce formalistic loyalty. This is not a neutral assumption, it is one made from a point of view colored by experiences with the law as a benign facilitator rather than an obstruction or tool of persecution. Lawyers are expected to act with conviction, but not have actual conviction. This leaves only one active word—lawyers must “act.” Thus, lawyers are expected not only to be advocates, but actors that give convincing imitations of loyalty and candor without conviction. This undermines the profession’s stature and credibility in the public eye. By adopting a strong version of neutral partisanship, the Model Code created a perverse system where only clients who are themselves “amoral” (either through role differentiation or an antisocial tendency to aver moral norms) gain the full benefit of an open relationship with their lawyers and full effective representation.

3. Clients and the Knowledge Deficit

Even if the Model Code managed to emulate loyalty effectively, neutral partisanship obscures a lawyer’s actual agenda and views, again disadvantage clients. By allowing lawyers to claim moral distinction from their client’s ends, lawyers inhabit a professional iden-
tity steeped in ambiguity. The client ought to be able to choose counsel knowing what her lawyer actually thinks of the moral content of the suit at hand. In a value-pluralistic society, it is likely that a client would find relevant what his lawyer really thinks of the moral validity of his case. However, the current system forces clients to make choices regarding their representation without knowledge of a lawyer’s moral view on the substantive issues. The client is required to trust that a lawyer’s personal views will not undermine the representation. This impedes, rather than empowers, client autonomy.

The client deserves, and arguably needs, to know what her lawyer believes. It is often vital for clients to know whether or not the lawyer is truly invested in the client’s legal goal or only performs a series of mechanically required acts as if she cared about the client. With such information at hand, the client, not the lawyer, can decide whether or not she is comfortable with a lawyer who does or does not share her moral views. This decision ought to lie with the client, at least under a view privileging client autonomy. Right now, only the lawyer knows of this potential disconnect, and the standard conception of lawyering paternalistically decides that this information is not relevant to the client (conveniently allowing an attorney to take on clients with conflicting moral positions for her own financial gain). Rather than allowing a lawyer to determine her impartiality (the fox guarding the henhouse), the rules could require that a lawyer reveal to the client any moral viewpoints on the representation she may have, allowing the client to consider these issues and waive any conflicts should the client feel that the lawyer, nonetheless, would be an excellent advocate.

Under a neutral-partisan system, the service that some clients receive will also be systematically better than that of others. The parties best served by the profession’s allegiance to the concept of role morality are those clients who are also disciples of the “neutral partisan” ethos. Clients who share the morally-charged view that parties or entities may separate professional activities from moral accountability are given the privilege (and advantage) of having a lawyer with genuinely aligned interests. But shouldn’t ordinary individual cli-

---

166. It is never possible to have complete knowledge in a given representation, even on the part of a lawyer. See, e.g., Spaulding, supra note 11, at 1384 (“[L]awyers know that their advice and advocacy will almost never be based on anything approaching complete knowledge.”). However, calculated and sanctioned failure to disclose information is a different matter.

167. Anne E. Thar, What Do Clients Really Want? It’s Time You Found Out, 87 ILL. B.J. 331 (1999) (reporting that more than skill or knowledge, what clients wanted from a lawyer was to know that the lawyer cared).
ents, unlikely to live in a role-differentiated manner, also have access to the same aligned, zealous representation?

Finally, a principal concern in the amoral lawyering justification of neutral partisanship is that lawyers will infringe on a client’s autonomy because of their disproportionate share of power in the relationship. On the basis of this perceived power-differential, neutral partisanship limits the power of lawyers and expands relative client power. While this may have been a genuine concern in the 1960s when the bar itself was filled with wealthy, privileged, predominately white, and generally elite individuals, as the project of diversifying and expanding the bar changed the composition of the bar, this reality may have also changed. The amoral lawyer defense of the neutral-partisan model assumes that (1) clients are less sophisticated than lawyers and (2) that the lawyers are financially independent and therefore are not subject to the client’s power of the purse.

Today, neither of these points is clear, particularly in the private sector. Many clients, particularly corporate entities, have more power in both society and in the client-lawyer relationship than their lawyers. Many, if not most, lawyers are not independently wealthy and many are saddled with debt. They are not, by and large, practicing the law solely as a vocation. It is a job and they must make money to live. Therefore, because clients control the money and can withdraw their business and go elsewhere, clients, not lawyers, often have the coercive power in the attorney-client relationship.

D. Defining Neutrality and the Limits of Self: Diversity and Equality

The final modern justifications for neutral partisanship are that it protects diversity and equality. However, this attempt is premised on accepting that: (1) neutrality protects those with minority viewpoints from the majority imposing its moral notions and (2) “neutral” rules lack a viewpoint and are impartial. This section questions these assumptions and looks specifically at how they interact with concepts of diversity and equality.

168. See Pepper, supra note 11, at 634 (“Because of the large advantages over the client built into the lawyer’s professional role, and because of the disadvantages and vulnerability built into the client’s role, the professional must subordinate his interest to the client’s.”).

169. See id.


171. See Pepper, supra note 140.
1. No One Is Neutral: An Example

The existence of human neutrality in any form is questionable; no person exists without a viewpoint.\textsuperscript{172} Rather, what would neutral qualities of lawyering look like without a background or viewpoint to fill that term with meaning? How can one define diligence or competence without a full human identity, without a background upon which to define such terms? Devoid of a set of experiences that shape her consciousness, a person would be less than human.\textsuperscript{173}

Take Professor Pepper’s article on neutrality as a case study of imbedded viewpoint specificity; here, Pepper’s experiences regarding social structures weaken his ideological arguments regarding the moral worth of individual autonomy and diversity. After raising arguments in favor of the moral worth of amoral lawyering on the basis of protecting unfettered client autonomy, Professor Pepper explains the concern that “many clients will come through the door without much internal moral guidance.”\textsuperscript{174} According to Professor Pepper, clients’ lack of moral guidance is the product of a “secularized society” that lacks “homogenous moral communities.”\textsuperscript{175} First, this assumes that because a lawyer fails to discern an internal guidance, it is in fact not there. Second, equating morality with immersion in a homogenous moral community is in tension with a commitment to valuing diversity. This statement seems to indicate that pluralism functions as an intermediary step towards a homogenous ideal: a benevolent assimilation process into a brave new homogenous (moral) world. And yet, it is the moral worth of pluralism that buttresses Professor Pepper’s argument for much of his article. Many “Americans take their pluralism as a fact to celebrate rather than a problem to be overcome”; however, here the lack of a singular moral vision in a community is viewed as a lack of moral substance.\textsuperscript{176}

The article goes on to reveal other imbedded assumptions regarding society and its norms and values. In discussing client morality, it paints a nostalgic picture of the past when clients gained their moral


\textsuperscript{173} See, e.g., ALASDAIR MACINTYRE, AFTER VIRTUE: A STUDY IN MORAL THEORY (2d ed. 1984) (arguing that the self is constituted by a life story with a purpose or “telos”).

\textsuperscript{174} Pepper, supra note 11, at 627. It is not without a certain irony that one can imagine that certain clients would feel similarly about the “neutral” lawyers that come through the door seemingly devoid of morality.

\textsuperscript{175} \textit{Id.}

\textsuperscript{176} Spaulding, supra note 11, at 1389 (discussing how in a value-pluralistic society referencing “ordinary morality” offered little substantive mooring).
footing from “[t]he rural town, the ethnic neighborhood, the church attended for several generations, the local business or trade community.”177 This is a world far removed from the “urbanized society” of the Model Code.178 Here the viewpoint provides a key insight: Professor Pepper’s description is ideal to some, while morally oppressive to others. Some may think of rural towns as quiet places where children play safely in the streets, but to others these same neighborhoods and groups were strongholds of provincialism, places where racially restrictive covenants, restrictive zoning, or community norms prevented many people from living productive lives. Many small trade groups (including the nascent ABA) historically excluded minorities and women. However, Professor Pepper’s experiences with these institutions may have led him to be less skeptical of these institutions as sources of morality, despite the fact that other rational thoughtful people may view them as the opposite based on their factual experiences.

Finally, when Professor Pepper asks, “would there have been more social justice, equality, or general welfare if lawyers had altered or withheld services on the basis of their own (largely middle- or upper-class) values?” he reveals not only that he assumes that lawyers are of a certain economic class but that the class to which they belong has some sort of unitary overarching set of beliefs.179 The answer to his question as phrased may be “no,” but this claim is falsifiable. There is also no reason why the socio-economic class of lawyers need be grounded in a particular narrow set of class values—unless the bar is failing to include people from varied backgrounds. The problem is not that lawyers will screen client values. The problem is that lawyers are a monolithic privileged group legislating over clients with whom they share little or no experience or values. How would we answer the following question: Would there have been more social justice, equality, or general welfare if lawyers from all walks of life represented clients whose legal ends they believed in?

The point here is not to denigrate Professor Pepper’s important scholarship and considerable influence. Rather, it is to point out that even a careful, well-intentioned, intelligent person who is trying to protect diversity, equality, and individual autonomy is unable to set aside the lens through which they interpret the world.

177. Pepper, supra note 11, at 627.
179. Pepper, supra note 11, at 620.
2. *Racism Without Racists*\textsuperscript{180}

Codifying neutral partisanship in the 1960s is not only problematic because the possibility of human neutrality is generally suspect. It is also an issue because neutrality had a particular socially charged understanding at that time. Neutrality, like color-blindness, was an attempt to conceptualize means for achieving social equality in a country of racially and ethnically diverse citizens. This nascent theoretical understanding of how to facilitate equal access to justice and free exercise of rights and how to strive towards equality shaped the Model Code (and by extension, the subsequent rules of conduct that have followed). While the concepts of the Model Code have remained relatively static in Model Rules of Professional Conduct today, understandings of inequality in a pluralistic society have developed further. This section seeks to bridge that divide.

Most likely, the drafters of the Model Code were familiar with arguments in favor of remedying social inequality through “colorblindness.” The concept of colorblindness was prevalent in discussions contemporary to the drafting of the Model Code as related to civil rights struggles of the 1950s and 1960s.\textsuperscript{181} The president of the ABA who called the Wright Committee, Justice Powell, went onto join the Supreme Court where he authored several important opinions regarding race and colorblindness.\textsuperscript{182} No doubt there is an appeal to the idea that the best way to make sure different people are treated fairly is to treat them all the same. The problem lies in determining how you will treat them all the same—and who defines that sameness.

The dangers of institutional mandates for neutrality are very similar to those imbedded in colorblindness in that both support, whether consciously or subconsciously, a hierarchy where the dominant groups’ status is the “norm” and all other groups are the non-neutral

\textsuperscript{180} I have borrowed this phrase from Eduardo Bonilla-Silva’s book of the same name. *EDUARDO BONILLA-SILVA, RACISM WITHOUT RACISTS* (2010) (arguing that colorblindness creates its own form of racism). *But see generally* GEORGE M. FREDRICKSON, *WHITE SUPREMACY* (1981) (arguing that racism should only apply to practices that hinge on white supremacy, the idea that whites as a race are inherently better and more capable than other races).


\textsuperscript{182} *See* McCleskey v. Kemp, 481 U.S. 279, 327 (1987) (finding no discrimination in Georgia’s death penalty system despite the fact that Georgia imposed death sentences on blacks who murder whites at twenty-two times the rate of blacks who murder blacks); Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265 (1978) (finding that racial and ethnic classifications of any sort are inherently suspect and call for strict scrutiny, even when remedial in nature).
“other.” Norms and assumptions impact how rules are written and how institutions function. This skew the system towards the preexisting status quo, regardless of intent, and creates a fissure between “our public commitments and our lived realities.” Under this understanding, outsider groups are allowed entrance into traditionally exclusive groups, such as the bar, only if they conform their conduct to the dominant groups’ norms. Successful assimilators gain access to the dominant group’s power. The price is that in exchange, these people’s distinct views, skills, and cultural influences are cast aside. Neutrality accommodates difference only by ignoring and therefore sublimating it. It does so by creating a neutral baseline comfortable to the dominant class of lawyers and based on their assumptions of what is neutral lawyer behavior. But assumptions are defined by this class’s experience and therefore are, necessarily, non-neutral.

Ian Haney-López lays out the limitations of colorblindness, noting that while it may have undercut ideas of white supremacy during the first phase of the civil rights movement, colorblindness was ineffective at dealing with systemic lingering racism. So too here, one can imagine that perhaps at the outset of the Model Code, there may have been utility to shielding lawyers for the moral pronouncement of the majority. But whatever efficacy that policy may have had in relation to overt recognition, it has run its course.

---

183. In this way, neutral partisanship falls prey to similar pitfalls to John Rawls’s original position. See JOHN RAWLS, A THEORY OF JUSTICE 136-42 (1971) (asserting as his original position that the best way to create a just society was to organize it from behind a “veil of ignorance” as if one did not have any specific identity or moral standpoint and also had no idea where one would be placed within the social order). Presumably this situation would allow the party involved to be impartial and fair. However, critiques of Rawls have clarified how in the attempt to create an impartial self, Rawls has actually created an individual who cannot exist as a real person, because a real person is part of incorporated entities, like the family, that have their own imbedded structure. See OKIN, supra note 172, at 96-99.

184. See, e.g., IAN HANEY LÓPEZ, WHITE BY LAW: THE LEGAL CONSTRUCTION OF RACE 157 (2006) (“Colorblindness is in this sense not a prescription but an ideology, a set of understandings that delimits how people comprehend, rationalize and act in the world.”).

185. Id. at 144.


187. See LÓPEZ, supra note 184, at 157-58 (“In the wake of the civil rights movement’s limited but significant triumphs, the relationship between colorblindness and racial reform changed remarkably. Whereas colorblindness in the context of Jim Crow was heavy with emancipatory promise, in the civil rights era and since, its greatest potency instead lies in preserving the racial status quo.”).

188. See Neil Gotanda, A Critique of “Our Constitution is Colorblind,” 44 STAN. L. REV. 1, 6 (1991) (“[B]efore a private person or a government agent can decide ‘not to consider race,’ he must first recognize it. In other words, we could say that one ‘noticed race but did not consider it.’ “).
The issue we face now is one of structural inequality. Colorblindness still requires that whoever enters the dominant group assimilate or “pass” as white:

We must be careful not to discount the willingness of significant sectors within the White community to extend a presumption of full human worth to racial minorities—nor should we be surprised that this presumption of full humanity often translates into treating ostensibly non-White persons as if they were White.190

Likewise, in the legal profession, compliance with neutral partisanship awards women and minorities full lawyer status for conforming their behavior to the preexisting norms established in lawyering prior to their entrance into the profession. As such, being treated “like a lawyer” translates into non-male, non-white, non-traditional lawyers being treated as the preexisting demographic of the bar (white men), or, alternatively, the way well-meaning white men might paternalistically treat women and minorities.

3. The Model Code Decoded: Examples of Non-Neutral “Neutrality”

Doctrinally, neutral partisanship is supported by attempting to set a baseline of conduct that is unpolticized. However, the very conceptions of what is a lawyer’s proper role, what are a lawyer’s mandatory duties, and how daily practice unfolds have rippling repercussions for the bar generally, certain lawyers in particular, disenfranchised clients, and the legal system. These choices are not neutral, nor can they be. The choices made in the Model Code, and now the Model Rules, reflect a certain view of agency and the lawyer’s prioritization of duties to client, law, society, and broader morality. It is a moral view, and it cannot be neutral. In it, agency is defined in terms of a certain type of loyalty, a certain type of candor (regarding the law but not the lawyer’s self), a certain type of competency, and confidentiality. In it, duties to clients are paramount. The examples provided below are meant to be illustrative of this point, but not exhaustive.

The 1969 Model Code set priorities for the legal profession by enumerating skills or duties necessary for lawyering in the disciplinary rules. These enumerated duties included competence, diligence, loyalty, and candor.191 In keeping with established conventions of

190. LÓPEZ, supra note 184, at 154.
code interpretation, anything not on the list is excluded.\textsuperscript{192} Thus, the Model Code's explicit articulation of relevant skills also creates a list of irrelevant skills. Lawyers who have strengths that are not neutral under status quo-defined baseline are at a disadvantage.\textsuperscript{193}

For example, in attempting to reach a neutral baseline, the Model Code adopted a version of lawyering that minimized and marginalized skills associated with women, particularly skills emphasizing emotional intelligence or group consciousness.\textsuperscript{194} Why? The Model Code was the first time the bar attempted to write rules that defined a baseline of conduct in a “neutral” way for a pluralistic group. The Canons, in contrast, did not need to be “neutral” across differences—the bar was generally homogenous. In the attempt to create a neutral code, the drafters of the Model Code most likely excised qualities that they felt were not neutral. However, since they defined a baseline from their viewpoints, a neutral code was one without “gendered” qualities—meaning those qualities associated with women.

Some of the qualities that could be included on this list are compassion, empathy/sympathy, and relationship building.\textsuperscript{195} Compassion is not only excluded as a necessary lawyerly trait: one could argue it is expressly barred. For example, DR-103(B) forbids the ad-

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{192} In statutory interpretation terms this is known as “\textit{expressio unius est exclusio alterius},” meaning “the express mention of one thing excludes all others.” BLACK'S LAW DICTIONARY (9th ed. 2009).
\item \textsuperscript{193} The exclusion of these traits has actually led to a norm that is generally harmful to the profession, as the public has grown to perceive lawyers as heartless and self-interested, and lawyers themselves have fallen into higher rates of depression and dissatisfaction. See Jacquelyn Smith, The Happiest and Unhappiest Jobs in America, FORBES (Mar. 22, 2013, 2:55 PM), www.forbes.com/sites/jacquelynsmith/2013/03/22/the-happiest-and-unhappiest-jobs-in-america/ (listing associate attorney as the most unhappy profession in America).
\item \textsuperscript{194} For the sake of demonstration, we will use here the broad brush stroke definition of gendered qualities articulated by different voice feminism: men are associated with rules, enumerating duties, and individual ethics, while feminine qualities including sensitivity to others, loyalty, responsibility, self-sacrifice, and peacemaking all reflect interpersonal involvement. See generally CAROL GILLIGAN, IN A DIFFERENT VOICE: PSYCHOLOGICAL THEORY AND WOMEN'S DEVELOPMENT (1982). While this breakdown is simplistic, see Angelina P. Harris, Race and Essentialism in Feminist Legal Theory, 42 STAN. L. REV. 581 (1980), it provides a broad jumping-off point that continues to be echoed in popular discourse to this day. See also LOUANN BRIZENDINE, THE FEMALE BRAIN (2006) (arguing that women’s brains are configured to render them more adept to social and group interactions).
\item \textsuperscript{195} In relation to civility as well, the Canons make a stronger case than the Model Code. Specifically, the Canons discuss the need for a “respectful attitude” and expressly frown on using lawsuits to obtain funds from clients. See Final Report, supra note 18, at 579 (Canons 1, 14). Lawyers are chastised for engaging in “unseemly wrangling” as “[a]ll personalities between counsel should be scrupulously avoided.” Id. at 580 (Canon 17). In questioning witnesses, the Canons caution against “improper speech” and “offensive personalities.” Id. at 580 (Canon 18); see also MODEL CODE OF PROF'L RESPONSIBILITY DR 7-101 (1980) (requiring only generally at (A)(1) that lawyers “avoid[] offensive tactics, [and] treat[] with courtesy and consideration all persons involved in the legal process”).
\end{itemize}
\end{footnotesize}
vancement of funds to clients beyond specific circumstances relating to supporting the litigation. At no point is a lawyer permitted to give funds to a client outright, regardless of need. The Model Code and the current Model Rules fail to recognize a duty to provide emergency aid or need. In fact, the Model Code does not include an overall duty of a lawyer to communicate effectively with her client, although one rule requires that the client is notified of the receipt of assets, and there are several rules that regulate communications with those with adverse interests as well as contact with witnesses, investigators, and the press.

Communication in general is a skill which women are often viewed as excelling in. While today a “duty to inform” a client is included in the Model Rules, this initial omission is noteworthy.

Perhaps more telling than what was included in the Model Code from the Canons is what was removed or downplayed. Specifically, the Model Code deemphasized community building and relationships at the bar, emotional consciousness, and civility, qualities easily associated with traditional “feminine” qualities, despite the fact that these qualities are valued by clients.

The exclusion of certain skills that were previously present in the Canons is particularly insightful given the Code generally greatly expanded, rather than limited, the Canons. The preface of the Model Code notes a main goal behind drafting new codes was to add “important areas of conduct that were either only partially covered or totally omitted from the Canons” and observed that the Canons were “sound in substance” but needed editorial revision.

As an initial matter, the Canons take a stronger view of the interconnected nature of the legal community and the duties of collegiality. The Canons expect lawyers to have loyalty to one another and an

---

196. See Model Code of Prof’l Responsibility DR 5-103(B), EC 5-8 (1980).
197. See id.
198. See id. at DR 9-102(B)(1) (“A lawyer shall: Promptly notify a client of the receipt of his funds, securities, or other properties.”).
199. See id. at DR 7-104, DR 7-107-10.
201. Model Rules of Prof’l Conduct R. 1.4 (requiring under the title “communication” in part B that “[a] lawyer shall explain a matter to the extent reasonably necessary to permit the client to make informed decisions regarding the representation”).
202. Ann Juergens, Valuing Small Firm and Solo Law Practice: Models for Expanding Service to Middle-Income Clients, 39 WM. Mitchell L. Rev. 80, 113 (2012) (reporting results of a study of small firms and solo practitioners who served moderate-income clients, which found that the principal factors in lawyer success were relationship building, communication, and collaboration with non-lawyers).
203. The Canons were originally only nine pages long, while the Model Code was over fifty pages long when it was first passed.
overall sense of civility. For example, Canon II (7) states that even if a client would like assistance of additional counsel, “[a] lawyer should decline association as colleague if it is objectionable to the original counsel.”\textsuperscript{205} The Canons also outline how “[e]fforts, direct or indirect, in any way to encroach upon the business of another lawyer are unworthy of those who should be brethren at the Bar.”\textsuperscript{206} However the Code makes no mention of such considerations nor of a duty of civility even in the broadest terms.

In compiling the necessary skills of lawyers, the Model Code also omits empathy and sympathy despite the fact that the Canons themselves outline specific instances mandating emotionally conscious actions. For example, the Canons admonish that an attorney should refrain from pushing trial forward where opposing counsel “is under affliction or bereavement.”\textsuperscript{207} The Canons also barred attempts to “curry favor” with jurors through “fawning, flattery, or pretend solicitude.”\textsuperscript{208} In the role of the advocate as articulated in the Model Code, there is no requirement to communicate or advise, taking into account the fear, anxiety, or general holistic state of one’s client. This has ongoing repercussions. The bar’s reticence to recognize as valid lawyer skills emotional intelligence and interpersonal aptitude has intensified over time as rules related to compassion have tightened\textsuperscript{209} and even time-honored lawyerly traits with emotional resonance, such as zeal, have fallen into disfavor.\textsuperscript{210}

Notably, the Model Code not only omitted certain qualities from the discussion of legal ethics—it also created new explicit duties. For the first time, a free standing duty of competence was articulated by designating a section of the Model Code under Canon 6 “A Lawyer Should Represent a Client Competently.”\textsuperscript{211} Competence is also referenced in the first Canon of the Model Code, which states that “[a] lawyer should assist in maintaining the integrity and competence of the legal profession.”\textsuperscript{212} In the enforceable and binding disciplinary rules, the Model Code commanded as follows:

\begin{itemize}
\item \textsuperscript{205} Final Report, supra note 18, at 577 (Canon 7).
\item \textsuperscript{206} Id.
\item \textsuperscript{207} Id. at 581 (Canon 24).
\item \textsuperscript{208} Id. at 581.
\item \textsuperscript{209} This element of the Code continues to this day in Model Rule 1.8, which subjects lawyers to discipline even if they provide money to clients for necessary, emergency, or vital non-legal expenses. MODEL RULES OF PROF'L CONDUCT R. 1.8 (2002).
\item \textsuperscript{210} For a more thorough discussion of the removal of zeal as a requirement over time, see Anita Bernstein, The Zeal Shortage, 34 HOFSTRA L. REV. 1165 (2006); Lawrence J. Vilardo & Vincent E. Doyle III, Where Did the Zeal Go?, 58 LITIG. (No. 1) 1, 4 (2011).
\item \textsuperscript{211} MODEL CODE OF PROF'L RESPONSIBILITY Canon 6 (1980).
\item \textsuperscript{212} Id. at Canon 1.
\end{itemize}
A lawyer shall not: (1) Handle a legal matter which he knows or should know that he is not competent to handle, without associating with him a lawyer who is competent to handle it. (2) Handle a legal matter without preparation adequate in the circumstances. (3) Neglect a legal matter entrusted to him.213

Competence is defined in terms of substantive legal knowledge, preparation, and organization. It limited the ability of lawyers to take on cases in opposition to the existing bar, as practice in areas of law that were new to an attorney required an existing member of the bar to “associate[e]” with the work. Generally, competence was not assessed by clients, peers, or by senior attorneys in a mentorship style environment. Instead, the competence revolution manifested in the first multistate bar exams in 1972, the requirement of attendance at ABA accredited schools in most jurisdictions, compulsory bar membership with ongoing fees, and ongoing legal education requirements. Competence did not include emotional intelligence, negotiation skills, substantive knowledge of the community being served, communication skills, or demonstrations of substantive writing skills.

The Wright Committee made clear that competence was added to the Model Code as a measure for screening lawyers, arguing that “standards must be established to exclude people from admission to the bar who could not, or are not likely to, serve clients capably and well”214. The inclusion of a competence duty as defined and implemented erected a barrier to entry to the legal profession and a means to remove bar licensing. In particular, competence was highlighted as being of central importance to all parties, “only if all persons have access to the law, which requires that they have access to lawyers of integrity and competence,” will government be that “of laws, not of men.”215 This distinction is particularly meaningful given the historic context of the Code’s drafting, when the issue of adherence to law had very real consequences in terms of support or lack of support for acts of civil disobedience. Committee Chairman Wright also articulated a more self-serving rationale: that misconduct by individuals would reflect badly on all lawyers and the judicial system.216

213. Id. at DR 6-101(A)(1-3).
215. Id. at 7.
216. See id. at 8 (“Lawyers are the face of our legal system that laymen most often see; the impressions that laymen have of our legal system are often in large measure by their impressions of lawyers.”).
4. Day to Day Practice: Limited Lawyer’s Roles, Limited Lawyers and Their Clients

Divesting lawyers of broader social and moral responsibility beyond client service limits the lawyer’s role in playing a unique and influential part in civil society. This limited role of lawyers may disproportionately impact socially disenfranchised individuals who become lawyers. The divestment of broader societal power from the attorney’s role generally has a particularly acute impact for attorneys who do not have other means to access influence. These lawyers must use their education and professional status to impact causes they favor, as opposed to lawyers who are wealthy, well connected, or have other avenues for gaining influence. The limitation of the lawyer’s role may also mean that some of the most educated and privileged members of a minority group (those who have law degrees) may be substantially limited in their ability to act politically through their work.

The neutral-partisan dominant ethos may discourage lawyers who are politically or morally motivated to become lawyers from joining the bar, since the neutral-partisan model of lawyering does not expect them to act upon their convictions. Likewise, it may also increase in the relative strength of the amoral lawyering contingent of the bar skews the legal profession away from even discretionary exercises of lawyer moral autonomy.217 To a certain extent, the bar is self-selected: people who find the neutral-partisan ideal appealing are attracted to the practice of law. Likewise, the inverse is also true; those attracted to the practice of law to engage with broader issues of justice and morality are likely to be repelled. When these people consider or attend law school the “dominant view” sends the message that such views of a lawyer’s role are fringe and bordering on unprofessional and that neutral-partisan client service is the central and perhaps only legitimate goal of lawyering.

The neutral-partisan limited role of lawyers also quiets the voices of lawyers who serve disenfranchised clients. The dominant ethos left

217. In the discretionary spirit of the work of Bill Simon, the current Model Rules do allow for certain discretionary exercises of morality but do not compel lawyers to act under these circumstances. See MODEL RULES OF PROF'L CONDUCT R. 1.6(b) (2013) (providing for disclosures pursuant to all exceptions to attorney duty of confidentiality, including those related to serious harms to third parties and those that are discretionary); id. at R. 1.16(b) (providing that lawyers may, under the right circumstances, withdraw from the representation where they have serious moral objections); id. at R. 2.1 (providing that lawyers may advise client on extra-legal concerns including “moral economic, social and political factors” relevant to a client’s case); William H. Simon, Ethical Discretion in Lawyering, 101 HARV. L. REV. 1083 (1988). However, the discretionary exercise of lawyer morality does not eliminate imbedded issues regarding ambiguity and pervasive norms of practice as well as the toll that the neutral-partisan baseline takes on both lawyers and the public perception of lawyers.
in place under a standard conception regime is one that limits lawyer autonomy to a very narrow bandwidth—essentially rendering a lawyer’s legitimate exercise of power in society as coterminous with that of clients. Therefore, lawyers who have powerful and sophisticated clients will also have broad power in society and leeway to pursue long-term legal goals and strategies. As repeat players, clients who are already sophisticated enough to comprehend and utilize the power of the law for societal influence can seek incremental structural change that is to their benefit. However, these clients do not need neutral partisanship to protect their autonomy; they are informed and have ample means to use market forces to curtail and shape lawyer behavior. On the other hand, lawyers with disenfranchised clients are comparatively limited. Clients from disenfranchised groups may not have the material means to set aside immediate relief in favor of developing long-term interests. Therefore, a system which does not set lawyer moral integration as a baseline leaves broad civic and moral issues almost exclusively to those who already have considerable social influence by virtue of their legal sophistication and wealth.

5. Discussion, Dissent, and Homogeneity: The Value of Pluralism

Somewhat ironically, amoral lawyering seeks to facilitate individual rights, equality, and diversity in American society by attempting to eradicate a plurality of views at the bar, leaving a neutral conduit for clients’ pluralistic views/agendas. There is a certain democratization element here, if one believes, as was the case in the mid-sixties, that clients were the common people and lawyers represented the elite. However, the bar is larger and much more diversified today. In many lawyer-client relationships the client is more “elite” than the lawyer. In this context, a system in a democratic society that seeks to impose a false sense of homogeneity for the purpose of avoiding ideological conflict is fundamentally misguided. Ideological conflict and dialogue across difference is necessary and important to a functioning democracy. It is not to be avoided or quashed, particularly in a forum, such as law practice, with a strong pedagogical purpose and highly visible role in society.

Neutral partisanship buys at a heavy price the peaceable coexistence of lawyers and clients with whom they have a moral disconnect. To the extent that pluralism has value in American society, it must actually exist in the public space and be predicated on open difference, not concealment. Disagreement, discomfort, and discourse are necessary and beneficial to the growth of American democracy and to the development of just laws. Because of this, attempting to eradicate difference through “neutrality” may also be immoral. If the bar truly seeks to protect diversity and equality (moral goals according to the
modern apostles of neutral partisans), then legal ethical guidance cannot seek an ideal where lawyers behave as though they were a homogenous group when they are not, and should not be. Rather, peaceful coexistence despite difference is wrong if it is predicated on fabricating and imposing homogeneity.

Morally valid pluralism requires disagreement, not avoidance. Disagreement, as in “factions,” is useful and necessary in a democracy to allow all parties a fair hearing, impede rash and unjust action, and make incremental steps towards the development of morally worthy law.\(^{218}\) This is as true in a civil action as in the voting process. The process of having civil discussions encourages competition of ideas and forces established thought to continually be reexamined, tested, and improved. Homogeneity allows ideas and power to stagnate and threatens the stability and legitimacy of a democracy.\(^{219}\) However, difference must be open, apparent, and discussed for its value to be realized.

Lawyers act as civic teachers to their clients, explaining how the American system of justice is supposed to work, what civic obligations are, and how to engage in civil discourse across differences.\(^{220}\) Clients can educate lawyers about how law is actually working on the ground and present their own moral views. Clients and lawyers, specifically, and society at large, benefit from lawyers and clients engaging in discussions of morality and the law.\(^{221}\) The current system abbreviates both the client’s moral input and the lawyer’s.\(^{222}\) By requiring moral dialogue before selecting and seeking an outcome, the profession sends the reassuring message to clients and attorneys that morality does have a place in the consideration of law, supporting the legitimacy of legal institutions. Lawyers can feel pride in working towards a cause they believe in, and for an institution that values forthright conversation. The attorney client relationship would flour-

\(^{218}\) The Federalist No. 51, at 291 (James Madison) (Clinton Rossiter ed., 1999) (stating that factions are a necessary means of protecting minority interests: “Different interests necessarily exist in different classes of citizens. If a majority be united by a common interest, the rights of the minority will be insecure.”).

\(^{219}\) Arguments in favor of the valor of difference play out more often in the literature related to market economies, where homogeneity is seen as stifling agency, innovation, and creativity.

\(^{220}\) See generally Bruce A. Green & Russell G. Pearce, “Public Service Must Begin at Home”: The Lawyer as Civics Teacher in Everyday Practice, 50 WM. & MARY L. REV. 1207, 1235-37 (2009) (arguing that, at a minimum, lawyers have a duty to discuss ideals of civic duty and virtue with clients).

\(^{221}\) See LUBAN, supra note 128, at 38 (arguing that the discovery of moral differences need not end the attorney-client relationship and proposing a different ideal of a “law practice in which the lawyer who disagrees with the morality or justice of a client’s ends does not simply terminate the relationship, but tries to influence the client for the better”).

\(^{222}\) See Pepper, supra note 11, at 630.
ish as lawyers have conversations with their clients that are not steeped in knowledge inequalities; both client and lawyer have equal claim to moral expertise (which may well be none).

Because lawyers are in a unique position to understand the complexity of the law, it is vitally important for our society to have lawyers interact and examine the law applying their ordinary moral compass. The only time this becomes a problem is if we assume, as Stephen Pepper does, that lawyers share a monolithic and elitist sense of values. Thus, the problem in this model of lawyering is not the application of ordinary values to legal representation or laws, but the application of a limited subgroup’s ordinary values to these enterprises. In this way, moral legitimacy is intimately intertwined with diversifying the legal profession. However, for the reasons discussed earlier, the system as it stands fails to encourage the inclusion of divergent moral dialogue.

A general culture of dissent and discussion among the bar and their clients enfranchises citizenry. It combats unnecessary lawyer paternalism and legitimates government institutions. The sense that different voices will be heard encourages a culture of debate, participation, and investment in the project of governance. Compelling role integration, rather than differentiation for lawyers, adds actual transparency to the social structure of society, the attorney-client relationship, and the legal profession as a whole.

V. Conclusion

This Article took on three key questions to understanding modern legal ethics and the current law regulating lawyers: (1) When did lawyers stop discussing professional conduct in the workplace in moral terms? (2) Why was that choice made? (3) What are the failings and current impacts of that choice? As to the first point, this Article argues that the advent of the 1969 Model Code marked a break in the marriage between the ideas of professional conduct and role integration. In explaining this shift, the Article draws on historical sources and textual analysis to make the argument that the choice to limit moral agency was one grounded in a pragmatic need to facilitate a certain view of pluralism. Finally, the remainder of the Article outlines how the advantages of neutral partisanship in terms of pluralism (as defined by its advocates) fail in many respects.

The Model Code and subsequent professional norms are vestiges of a charged historical moment, a moment in which the language of

223. Particularly now when law is increasingly voluminous and specialized, lawyers may be the only parties who know of the complex apparatus of law in a given area and its ramifications.
difference was nascent and theoretical understanding of diversity, equality, and individual rights was underdeveloped. In that context, even well intentioned parties could perceive a commitment to divorcing ethics from professional conduct and justifying neutral partisanship as moral, based on a flawed conception of how to facilitate equality and plurality.

Yet lawyers, as a profession, continue to plow forward without taking into account that historic moment, the tensions inherent in that moment, and the missteps taken in accordance with those early misconceptions of facilitating change, equality, and pluralism at the bar. In creating a code that incorporates latent bias, delegitimizes the discussion of broad ethical themes, and undermines lawyer individualism by adopting a homogenizing neutral-partisan ideal, the Model Code fundamentally undermined meaningful long-term pluralism, legal ethics, and the profession as a whole.

Today, the legal profession and modern legal ethics continue to struggle with the historic legacies built deep into the doctrinal and normative framework of the Model Code. Setting aside moral discourse among lawyers in service of demographic pluralism and adopting neutral partisanship was a poor trade, grounded in fundamental theoretical misconceptions of what neutrality and pluralism require. Adopting neutral partisanship as the “standard conception” of lawyering prevented, rather than enabled, the meaningful inclusion of the values of new entrants to the bar and solidified an ideal of lawyers that is dehumanizing. Neutral partisanship limits all lawyers’ ability to legitimately articulate their moral judgment, differing viewpoints, and unique experiences in the workplace.

The story here is not one of calculated malevolence. The architects of the Model Code were reasonable people drafting what appeared at the time to be a sensible plan to respond to a changing professional landscape. But these drafters were limited by the scope of their personal experiences and historical context. Their limitations need not be ours.
WORK MADE FOR HIRE—
ANALYZING THE MULTIFACTOR BALANCING TEST

RYAN VACCA* 

ABSTRACT

Authorship of copyrighted works is oftentimes controlled by the 1976 Copyright Act's work made for hire doctrine. This doctrine states that works created by employees within the scope of their employment result in the employer, not the employee, being the author and initial copyright owner. One key determination in this analysis is whether the hired party is an employee or independent contractor. In 1989, the U.S. Supreme Court, in CCNV v. Reid, set forth a list of factors to distinguish employees from independent contractors. Unfortunately, the Supreme Court did not give further guidance on how to balance these factors. Three years later, in 1992, the Court of Appeals for the Second Circuit decided Aymes v. Bonelli and noted that not all factors are equally weighted and that five of the factors would "be significant in virtually every situation." This analysis was supported by looking at all the work made for hire cases decided in the three year period since Reid—six cases in total. This Article expands in both scope and time what the Second Circuit did in Aymes and systematically analyzes how courts have utilized the factors in the twenty-five years since Reid. In particular, this study has identified the universe of cases where the courts have decided whether a hired party was an employee or independent contractor and uses the data from these cases to describe what factors appear to be the most and least important in reaching these conclusions. Based on the results of this study, this Article proposes a continuum of importance, which graphically illustrates the relative importance of each factor.

I. INTRODUCTION ................................................................. 198
II. AUTHORSHIP AND THE WORK MADE FOR HIRE DOCTRINE ................................................................. 201
   A. Pre-1909 Case Law .......................................................... 202
   B. The 1909 Act .................................................................... 204
   C. The 1976 Act—Legislative History and Statutory Text ......................................................................................... 206
   D. Different Interpretations of "Employee" Under the 1976 Act ................................................................. 209
      1. Right to Control Test ......................................................... 210
      2. Actual Control Test .......................................................... 211
      3. Agency Test .................................................................. 213
      4. Formal, Salaried Employee Test ........................................ 216
   E. CCNV v. Reid and the Multifactor Balancing Test ......................................................................................... 218
   F. Weighting Factors—The Aymes v. Bonelli Three-Year Study .............................................................................. 222
III. UPDATING AYMES—A COMPREHENSIVE MULTIFACTOR STUDY ................................................................. 223
   A. Methodology ..................................................................... 224
   B. Results ......................................................................... 225
IV. DISCUSSION ................................................................................................................................. 231
   A. Doctrinal Fit ..................................................................... 232
   B. Practical Implications .......................................................... 236

* Associate Professor of Law and Director of Faculty Research & Development, University of Akron School of Law. This Article was selected as the recipient of the University of Houston Law Center's Institute for Intellectual Property & Information Law's 2012-2013 Sponsored Scholarship Grant for the Legal Academy. The author gratefully thanks Lydia Loren, Tomás Gómez-Arostegui, Anna Laakmann, Bruce Boyden, Eric Johnson, Tracy Reilly, Barton Beebe, Molly McBurney, and the participants at the University of Kentucky Developing Ideas Workshop, the Fifth Annual Conference on Innovation and Communications Law, the Thirteenth Annual IP Scholars Conference, the Lewis & Clark IP in the Trees Workshop, and the Marquette Law's Junior Faculty Scholars WIP Conference for their contributions, the University of Akron School of Law and University of Houston Law Center for their generous research support, and Theresa Rakocy for her invaluable research assistance.
Authorship plays a central role in copyright law.\footnote{See U.S. CONST. art. I, § 8, cl. 8 ("Congress shall have Power . . . [t]o promote the Progress of Science and the useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." (emphasis added)); see also Peter Jaszi, Toward a Theory of Copyright: The Metamorphoses of "Authorship," 1991 DUKE L.J. 455, 455 (1991) ("[A]uthorship . . . is arguably the most central, and certainly the most resonant, of the foundational concepts associated with Anglo-American copyright doctrine.")} Conferring this title on people or entities bestows on them initial ownership of the copyright\footnote{17 U.S.C. § 201 (2012).} and the power to exploit the associated rights.\footnote{Id. § 106 ("[T]he owner of copyright under this title has the exclusive rights to do and to authorize any of the following[] . . . ."); id. § 106A(a)(1) ("[T]he author of a work of visual art—shall have the right[] . . . .").} Hence, it is no surprise that authorship is an oft-disputed issue between parties contesting the use of copyrighted works.\footnote{See, e.g., JustMed, Inc. v. Byce, 600 F.3d 1118 (9th Cir. 2010); Martha Graham Sch. & Dance Found., Inc. v. Martha Graham Ctr. of Contemporary Dance, Inc., 380 F.3d 624 (2d Cir. 2004); Numbers Licensing, L.L.C. v. bVisual USA, Inc., 643 F. Supp. 2d 1245 (E.D. Wash. 2009); Latimer v. Roaring Toyz, Inc., 574 F. Supp. 2d 1265 (M.D. Fla. 2008); Int'l Code Council, Inc. v. Nat'l Fire Prot. Ass'n, No. 02C5610, 2006 WL 850879 (N.D. Ill. Mar. 27, 2006); Brower v. Martin, 446 F. Supp. 2d 232 (S.D.N.Y. 2006).}

In many circumstances, determining authorship is easy, because the general rule is that “the author is the party who actually creates the work, that is, the person who translates an idea into a fixed, tangible expression . . . .”\footnote{Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 737 (1989).} The artist who paints a bowl of fruit in her home studio on the weekend to sell at a community art show is an easy example of the traditional notion of authorship. The glaring exception to this general rule is the “work made for hire” doctrine. In part, this doctrine provides that “a work prepared by an employee within the scope of his or her employment”\footnote{17 U.S.C. § 101 (defining work made for hire).} is consequently owned by the employer.\footnote{Id. § 201(b).} But the work made for hire doctrine does not simply
change initial ownership of the copyright. This doctrine has found its way into several other issues, such as modifying the duration of the copyright, eliminating the right to terminate transfers of copyright, and prohibiting the acquisition of moral rights.

But unlike Hamlet, where the question was “to be, or not to be,” for the work made for hire doctrine, the question is “employee or independent contractor?” Distinguishing between employees and independent contractors under the 1976 Copyright Act’s work made for hire doctrine is an essential problem courts and parties need to resolve before proceeding to analyze how ownership, duration, terminations of transfers, and moral rights are affected.

Twenty-five years ago, the U.S. Supreme Court attempted to give guidance to judges and parties involved in disputes over the employment status of hired individuals. In 1989, the Court issued its opinion in Community for Creative Non-Violence v. Reid and declared that distinguishing employees from independent contractors should be accomplished by using a multifactor balancing test. This test, which uses approximately a dozen factors, clarified what the proper test was but spawned questions about how those factors are balanced and which factors, if any, are the most important in the analysis.

Three years after Reid, the Court of Appeals for the Second Circuit decided Aymes v. Bonelli and noted that not all factors are equally weighted. The court in Aymes went further still and opined that five of the factors would “be significant in virtually every situation.”

---

9. Compare 17 U.S.C. § 302(a) (providing that copyright generally persists for the life of the author plus seventy years), with id. § 302(c) (providing that copyright in a work made for hire “endures for a term of 95 years from the year of its first publication, or a term of 120 years from the year of its creation, whichever expires first”).
10. Id. § 203(a) (“In the case of any work other than a work made for hire, the exclusive or nonexclusive grant of a transfer or license of copyright or of any right under a copyright[,] . . . is subject to termination . . . .”); see also id. § 304(c), (d) (same exclusion but applied to a different set of transfers).
11. See id. § 106A(a) (granting moral rights to “the author of a work of visual art”); id. § 101(2)(B) (defining “work of visual art” as not including a work made for hire).
15. Id.
16. Id. at 751-52.
17. See infra Parts II.E–F.
19. Id.
This analysis was supported by reviewing all the work made for hire cases decided in the three-year period since Reid—six cases in total.\footnote{20} This Article greatly expands upon the Second Circuit’s work in Aymes. It is the first study to comprehensively and systematically analyze the work made for hire factors and show their relative importance in distinguishing employees from independent contractors. It contributes to the existing and highly-regarded literature of empirical studies analyzing multifactor tests.\footnote{21} Using a data set consisting of the population of work made for hire cases decided since Reid where the court determined the status of the hired party, this Article illustrates which factors are the most and least important in drawing this distinction. These results confirm, in part, what the Second Circuit in Aymes believed—that a small subset of factors is important.\footnote{22} Likewise, these results show that some factors originally thought to be important are not terribly important and that other factors originally thought to be of only moderate importance are actually very important.\footnote{23}

Part II provides historical background on the work made for hire doctrine and illustrates how the doctrine has evolved from when its foundation was laid in the mid-1800s through the Second Circuit’s 1992 decision in Aymes.\footnote{24} Specifically, during the course of this discussion, I describe the legislative histories leading up to the codification of the work made for hire doctrine in the 1976 Copyright Act\footnote{25} and the circuit split occurring after its enactment, which eventually led to the Supreme Court’s resolution in Reid.\footnote{26} Part II concludes by recounting the Supreme Court’s rationale for adopting the multifactor test in Reid\footnote{27} and the Second Circuit’s explanation of the important and unimportant factors in Aymes.\footnote{28}

Part III describes the methodology and results of this study.\footnote{29} This Part explains the type of data collected, describes how this data is used to show four different measures of importance, and displays

\footnote{20}{See id. at 862 (citing six cases decided between 1990 and 1992).}


\footnote{22}{See infra Part II.F.}

\footnote{23}{See infra notes 260-61 and accompanying text.}

\footnote{24}{See infra Part II.}

\footnote{25}{See infra Parts II.A–C.}

\footnote{26}{See infra Part II.D.}

\footnote{27}{See infra Part II.E.}

\footnote{28}{See infra Part II.F.}

\footnote{29}{See infra Part III.}
how each factor ranks using each measure. Most importantly, based on the results of these measures, I propose a continuum of importance for the Reid factors. This continuum groups together factors sharing similar features with respect to the measures of importance. Part III concludes by testing how useful the continuum is at predicting the ultimate outcome of the work made for hire cases.

Part IV first attempts to situate the results of this study within the historical development of the work made for hire doctrine and see whether these results cohere with the Supreme Court’s rationales for adopting and rejecting the various interpretations existing before Reid. Part IV then considers the implications of this study, paying attention to litigation and business-planning contexts. Special attention is drawn toward terminations of transfers litigation involving sound recordings as a case study on how the results of this study may ultimately affect the outcome of this impending litigation.

II. AUTHORSHIP AND THE WORK MADE FOR HIRE DOCTRINE

To understand the modern work made for hire doctrine, it is necessary to appreciate how it came into being. Many of the changes taking place in 1976 and subsequent interpretations were heavily influenced by earlier copyright acts and how courts construed the work made for hire doctrine. The most significant of these early acts was the 1909 Copyright Act. This part of the Article briefly describes the work made for hire doctrine during the pre-1909 period and then discusses the codification of this doctrine in the 1909 Act and how the courts gradually, but greatly, modified this doctrine. Afterwards, this part describes the legislative history leading to the 1976 Act and the final codification in section 101, which left courts struggling with how to determine whether a hired party was an employee or independent contractor. Next, this part explains the four approaches lower courts used in making this determination before the Supreme Court resolved the issue in its historic 1989 case—CCNV v. Reid. Finally, it describes how the Second Circuit in Aymes v. Bonelli put an additional gloss on Reid, which serves as the jumping off point for this study and is fully explored in Parts III and IV.

30. See infra Part III.
31. See infra Fig. 1 in Part III.B.
32. See infra Fig. 1 in Part III.B.
33. See infra Fig. 2 and preceding discussion in Part III.B.
34. See infra Part IV.A.
35. See infra Part IV.B.
36. See infra Part IV.B.2.
A. Pre-1909 Case Law

The earliest appearance of the work made for hire doctrine (or at least the foundation for the doctrine) occurred in the 1860s. Prior to this time, hiring parties were never entitled to the copyright of the parties they hired absent an agreement assigning the copyright to the hiring party. As such, no distinction was drawn between employees and independent contractors, because hiring parties were not entitled to the copyright by virtue of employment.

During the 1860s, a shift occurred and courts began to recognize that hiring parties did have a copyright interest in the works prepared by those they hired to create. One of the earliest cases laying the foundation for the work made for hire doctrine was Keene v. Wheatley. Keene involved the famous play, Our American Cousin. In relevant part, the British author, Tom Taylor, sold the American rights to Laura Keene, a New York City theater owner. Joseph Jefferson, an actor in Keene’s company, adapted the play for performance in Keene’s New York theater. The play became a huge success in New York, and Jefferson sold his additions in the play to Wheatley and John Clark, two Philadelphia theater producers, who had previously acquired a British copy of the play. After a successful exhibition of the play in Philadelphia, Keene filed suit alleging copyright infringement. The court held that Keene did not have a valid copyright to the original version of the play, because Taylor was not a U.S. resident. Notwithstanding this ruling, the court also held that Keene could seek a remedy against Wheatley and John Clark for Jefferson’s additions to the play. As the court explained:

Mr. Jefferson, while in the general theatrical employment of the complainant, engaged in the particular office of assisting in the adaptation of this play; and made the additions in question in the course of his willing performance of this duty. She consequently

38. Id. at 32.
39. Id.
40. Id.
42. Id. at 181. Our American Cousin is famously known as the play occurring at Ford's Theater the night President Lincoln was assassinated by John Wilkes Booth. Fisk, supra note 37, at 38 n.125.
43. Keene, 14 F. Cas. at 182.
44. Id.
45. Fisk, supra note 37, at 37-38.
46. Keene, 14 F. Cas. at 184.
47. Id. at 185; Fisk, supra note 37, at 39-40.
48. Fisk, supra note 37, at 40.
In particular, the court declined to hold that Keene was entitled to the copyright as a matter of statutory copyright law but, instead, based its holding on equitable principles.\(^{49}\)

Eight years later, in *Lawrence v. Dana*,\(^{51}\) a federal court collapsed the distinction between statutory copyright law and equitable principles.\(^{52}\) *Lawrence* involved an international law treatise written by Henry Wheaton.\(^{53}\) Upon Henry Wheaton's death, his wife, Catharine, sought the assistance of William Lawrence to prepare a new edition of the treatise.\(^{54}\) As part of the agreement between Catharine and Lawrence, Catharine agreed to not make use of Lawrence's notes in a new edition of the treatise without his consent.\(^{55}\) Subsequent litigation over the treatise ensued, and in dictum the court indicated that had the parties not entered into a contract for Lawrence to retain the copyright, Catharine would have owned the copyright because of the employment relationship between them.\(^{56}\) This dictum was repeated throughout the rest of the nineteenth century as courts reallocated copyright ownership by including an implied contract between the parties that favored the hiring party owning the copyright.\(^{57}\) Despite courts entertaining the idea of the hiring party owning the copyright in the absence of an express assignment to that effect, there was confusion about whether the default rule favored the hiring or hired party.\(^{58}\) The cause for this shift from a default rule of hired party ownership to a default rule of hiring party ownership is muddled.\(^{59}\) Nonetheless, one thing is clear—the courts deciding copyright cases during this period drew no distinction between employees and independent contractors like is done in the modern context.\(^{60}\)

---

49. *Keene*, 14 F. Cas. at 187.
50. *Id.*, Fisk, *supra* note 37, at 40.
51. 15 F. Cas. 26 (C.C.D. Mass. 1869) (No. 8136).
52. Fisk, *supra* note 37, at 43.
54. *Id*.
55. *Id*.
56. *Id*. at 51; Fisk, *supra* note 37, at 43.
57. Fisk, *supra* note 37, at 45.
58. *Id*. at 47 (“By the last two decades of the nineteenth century, the law of employee copyrights was highly uncertain and the results of cases were quite unpredictable.”).
59. *Id*. at 45 (suggesting the cause of this shift was based on (1) reflecting the intent of most parties, (2) employers having a stronger moral claim, or (3) changing assumptions about the nature of authorship).
60. *Id*. at 46.
It was not until nearly the turn of the century that the principle of hiring party-owned copyrights was more firmly established. In 1899, the court in *Collier Engineer Co. v. United Correspondence Schools* noted that it was the employer of a salaried employee who was the copyright owner of instructional materials for a correspondence school. The following year, the court in *Dielman v. White* held that “when an artist is commissioned to execute a work of art not in existence at the time the commission is given, the burden of proving that he retains a copyright in the work of art executed, sold, and delivered under the commission rests heavily upon the artist himself.”

**B. The 1909 Act**

After an uncertain history, Congress finally recognized the work made for hire doctrine in the Copyright Act of 1909. The 1909 Act provided that “the word ‘author’ shall include an employer in the case of works made for hire.” This provision of the 1909 Act resulted from a meeting in 1905 between industry leaders and the American Authors’ League. The initial draft stated that only “authors” could obtain copyright, but industry leaders objected. The Copyright Office proposed a revised bill that did not have a general work made for hire provision, but instead had a section indicating that the “publisher of a composite or collective work . . . which has been produced at his instance and expense” would be entitled to a copyright. However, industry leaders were still unsatisfied. As one participant complained:

We have people who work for us who make engravings or etchings for us under salary. Under the new law—if it becomes a law as drafted—they would have the right to copyright, and I think it would be well to express in such a law that where no agreement exists to the contrary the payment of a salary to an employee shall entitle an employer to all rights to obtain a copyright in any work

---

61. *Id.* at 55.
63. *Fisk, supra* note 37, at 59-60.
65. *Id.* at 894.
68. *Id.*
69. *Id.*
70. *Id.* at 64 (citing 1 LEGISLATIVE HISTORY OF THE 1909 COPYRIGHT ACT xxiv (B. Fulton Brylawski & Abe Goldman eds., 1976)).
performed during the hours for which such salary is paid. It seems to me these things should not be left to the courts to decide.\textsuperscript{71}

Interestingly, this comment indicates that some industry leaders believed that the copyright in works created by salaried employees during their employment would belong to the employer.\textsuperscript{72} The Copyright Office was persuaded by such comments and prepared another draft of the bill, which credited authorship to “[a]n employer, in the case of a work produced by an employee during the hours for which his salary is paid, subject to any agreement to the contrary.”\textsuperscript{73} Later, in 1906, Congress held joint hearings and produced the language ultimately ending up in the 1909 Act—“the word ‘author’ shall include an employer in the case of works made for hire.”\textsuperscript{74} Unfortunately, Congress failed to define “work made for hire” and “employer.”\textsuperscript{75}

One of the earliest cases interpreting the 1909 Act’s work made for hire provision was National Cloak & Suit Co. v. Kaufman.\textsuperscript{76} In Kaufman, the plaintiff, a corporation, claimed a copyright in a fashion book.\textsuperscript{77} The defendant reproduced portions of this book and, when sued for copyright infringement, moved to dismiss.\textsuperscript{78} In the course of its opinion, the Second Circuit discussed the newly enacted 1909 Act and the work made for hire doctrine.\textsuperscript{79} Importantly, the court noted that “[under the previous law], as well as now, the employer had the right to the copyright in the literary product of a salaried employee.”\textsuperscript{80}

The last major development in the work made for hire doctrine before enactment of the 1976 Act was Brattleboro Publishing Co. v. Winmill Publishing Corp.\textsuperscript{81} In Brattleboro, newspaper advertisers hired the Brattleboro Daily Reformer, a local newspaper, to create advertisements.\textsuperscript{82} These advertisements appeared in the Brattleboro Daily Reformer and subsequently appeared in the Brattleboro Town Crier, a direct-mail circular.\textsuperscript{83} The Brattleboro Daily Reformer sued the Brattleboro Town Crier for copyright infringement for reproduc-

\begin{flushleft}
\textsuperscript{71} Id. (citing LEGISLATIVE HISTORY, supra note 70, at 65).
\textsuperscript{72} See id. (citing LEGISLATIVE HISTORY, supra note 70, at 65).
\textsuperscript{73} Id. at 66 (citing LEGISLATIVE HISTORY, supra note 70, at xxx).
\textsuperscript{74} Id. (citing LEGISLATIVE HISTORY, supra note 70, pt. J).
\textsuperscript{75} 1 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 5.03[B][1][a][i] (2014).
\textsuperscript{76} Nat’l Cloak & Suit Co. v. Kaufman, 189 F. 215 (C.C.M.D. Pa. 1911).
\textsuperscript{77} Id. at 216.
\textsuperscript{78} Id.
\textsuperscript{79} Id. at 217.
\textsuperscript{80} Id. (emphasis added).
\textsuperscript{81} Brattleboro Publ’g Co. v. Winmill Publ’g Corp., 369 F.2d 565 (2d Cir. 1966).
\textsuperscript{82} Id. at 568.
\textsuperscript{83} Id. at 567.
\end{flushleft}
ing the advertisements. On appeal, the Second Circuit stated that in determining whether a work was a work made for hire under the 1909 Act, it applied the “instance and expense” test. That is, the copyright is owned by “the person at whose instance and expense the work is done.” In Brattleboro, the court held that the advertisers, who hired the newspaper, were the copyright owners of the advertisements. Importantly, the Second Circuit opined that the “instance and expense” test applied regardless of whether the hired party was a traditional employee or an independent contractor. Collapsing the distinction between employees and independent contractors laid the foundation for some of the confusion arising from the 1976 Act's work made for hire provision.

C. The 1976 Act—Legislative History and Statutory Text

The deluge of technological advancements created in the wake of the 1909 Act caused the judges interpreting it to stretch the statutory language to its limits to accommodate these developments. As a result, in 1955, Congress decided to overhaul the 1909 Act, which included funding thirty-five studies on copyright issues. One of these studies, study number thirteen, published in 1958, focused on the work made for hire issue and reported the then-current state of the law under the 1909 Act.

Based on these studies, the Copyright Office prepared a report on copyright law revisions, including recommendations on how to deal with the work made for hire issues. This 1961 report recommended that works made for hire be defined as “works created by an employee within the regular scope of his employment” and suggested that commissioned works fell outside this definition. Following publica-
tion of the 1961 report, stakeholders voiced their concerns about the recommendations.95 One objection lodged by the motion picture industry concerned the use of the phrase “regular scope of employment” in the definition.96 The movie studios thought this language narrowed what the studios had typically been able to claim as their own.97 None of the objections concerned the understanding of what the term “employee” meant.98 In fact, as Professor Trotter Hardy describes it, the comments to the 1961 report implicitly recognized “employee” to mean formal employees—those paid on a regular schedule.99

After these discussions, the Copyright Office published a “Preliminary Draft Bill” in 1963.100 This bill defined work made for hire as “a work prepared by an employee within the scope of the duties of his employment, but not including a work made on special order or commission.”101 Although the 1963 Preliminary Draft Bill included a few changes to the work made for hire doctrine, it did not define the term “employee.”102 Stakeholders’ understanding of the term “employee” became clear as a result of the objections to the provision in the 1963 Preliminary Draft Bill stating that specially ordered or commissioned works were not works made for hire.103 As the stakeholders argued about the status of commissioned works, they frequently contrasted commissioned individuals with employees.104 Again, as Professor Hardy articulately describes, these references to employees almost always describe employees as individuals who were formal employees—those who were on the payroll and paid a regular salary.105

95. Hardy, supra note 90, at 225-28; Landau, supra note 66, at 115.
96. Hardy, supra note 90, at 225 n.66.
97. See id. Another objection to the proposal was that employers would not be considered authors, but merely owners of the copyright. Id. at 226.
99. Hardy, supra note 90, at 226-27.
100. Id. at 228; Landau, supra note 66, at 115.
101. Hardy, supra note 90, at 228 (emphasis added); see also Landau, supra note 66, at 115.
102. Hardy, supra note 90, at 229.
103. Id. at 232-35. The debate about commissioned works was that federal copyright law would preempt common law copyright, which had treated hiring parties that commissioned works as the copyright owner. In effect, commissioned works were treated similarly to employee-created works, but with preemption in effect that common law development would no longer exist and the hiring party would not own the copyright unless the hired party agreed to assign it. This, in conjunction with the introduction of the termination of transfer provisions, which did not apply to works made for hire, caused much consternation for film, textbook, and reference publishers. Id. at 229-32.
104. Id. at 232-35.
105. Id.
Following this set of discussions, Congress introduced the 1964 bill, which included commissioned works within the definition of “work made for hire.” In fact, the 1964 bill provided that any specially ordered or commissioned work would be a work made for hire “if the parties so agree[d] in writing.” This one hundred eighty degree change on specially commissioned works raised objections from individual creators who had been routinely hired as independent contractors, especially members of the Authors League. After much haggling, the two sides of the work made for hire debate were able to find some common ground and agreed that some types of specially commissioned works should be works made for hire, while other types were not appropriate for that treatment.

At last, in 1965, Congress introduced two bills which defined “work made for hire” as “a work prepared by an employee within the scope of his or her employment” and also specially ordered or commissioned works falling within four categories of works, including a contribution to a collective work, parts of a motion picture, translations, and supplementary works, if the parties expressly agreed in writing that the work shall be considered a work made for hire. After further discussions, this list of categories was expanded to include compilations, instructional texts, tests, atlases, and answer materials for a test.

106. Technically, Congress introduced three identical bills—S. 3008, H.R. 11947, and H.R. 12354. For the sake of simplicity I refer to them collectively as the “1964 bill.” See Hardy, supra note 90, at 236.

107. Landau, supra note 66, at 115.

108. Hardy, supra note 90, at 236; Landau, supra note 66, at 115.

109. Hardy, supra note 90, at 237. These individuals were concerned that they could easily be forced to sign a contract declaring that the work was a work made for hire and, as a result, their termination rights would never exist. Landau, supra note 66, at 115-16.

110. Hardy, supra note 90, at 238.

111. See generally 1963 REG. REP., supra note 98 (some commentators noting their concerns about including specially commissioned works resulting in too little protection for authors; some commentators noting that a substantial percentage of audio-visual materials are specially commissioned and that the proposed law needs to be changed).


113. Landau, supra note 66, at 116 (internal quotations omitted).

114. Hardy, supra note 90, at 238-40; Landau, supra note 66, at 116. One of the driving forces behind these debates about the work made for hire definition was the new termination of transfers provisions. See Landau, supra note 66, at 116 (“[I]n exchange for concessions from publishers on provisions relating to termination rights, the authors consented to a second subsection . . . .”).

115. Landau, supra note 66 at 117. The stakeholders involved in negotiating these categories—hence the seemingly random collection of included works—were representatives of the Authors League of America, the Ad Hoc Committee on Educational Institutions and Organizations, the American Guild of Authors and Composers, the American Textbook
focused on the specially commissioned works; but in reference to this issue, they explained their understanding of what an employee was.\textsuperscript{116} As they had in the past, the stakeholders considered employees as those who were paid a salary for either a fixed or indefinite term.\textsuperscript{117}

After more than a decade of debates about the work made for hire doctrine, the statutory language had been worked out and was enacted, albeit a decade later, in the 1976 Copyright Act.\textsuperscript{118} In relevant part, the definition of “work made for hire” now reads:

A “work made for hire” is—

(1) a work prepared by an employee within the scope of his or her employment; or

(2) a work specially ordered or commissioned for use as a contribution to a collective work, as a part of a motion picture or other audiovisual work, as a translation, as a supplementary work, as a compilation, as an instructional text, as a test, as answer material for a test, or as an atlas, if the parties expressly agree in a written instrument signed by them that the work shall be considered a work made for hire.\textsuperscript{119}

Although the 1976 Copyright Act uses the phrase “a work prepared by an employee within the scope of his or her employment,” it fails to define the term “employee.”\textsuperscript{120} Nonetheless, given the extensive legislative history, one possibility was that everyone understood the term to mean those who were formal, salaried employees.\textsuperscript{121}

\textbf{D. Different Interpretations of “Employee” Under the 1976 Act}

The 1976 Copyright Act went into effect on January 1, 1978, and courts were quickly confronted with the task of having to interpret the work made for hire language in the statute.\textsuperscript{122} During the course of the following eleven years, the lower courts adopted four different tests to determine whether a hired party was an employee under the first prong of the work made for hire definition.\textsuperscript{123}

\begin{footnotes}
\footnote{Hardy, \textit{supra} note 90, at 239.}
\footnote{\textit{Id.} (discussing the testimony of Adolph Schimel on behalf of the Motion Picture Association of America).}
\footnote{Landau, \textit{supra} note 66, at 117.}
\footnote{17 U.S.C. § 101 (2012) (defining work made for hire).}
\footnote{Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 738 (1989) (internal quotations omitted).}
\footnote{Hardy, \textit{supra} note 90, at 241; Litman, \textit{supra} note 89, at 890.}
\footnote{Marita Covarrubias, Note, \textit{The Supreme Court Sculpts a Definition . . . Is It a Work for Hire?}, 10 LOY. ENT. L.J. 353, 365 (1990).}
\footnote{Landau, \textit{supra} note 66, at 120-21.}
\end{footnotes}
1. Right to Control Test

The first test was known as the “right to control” test.\textsuperscript{124} Under this test, courts looked at “whether the alleged employer has the right to direct and supervise the manner in which the writer performs his work.”\textsuperscript{125} This test was borrowed from case law interpreting the 1909 Act.\textsuperscript{126}

One of the first cases adopting the right to control test was \textit{Town of Clarkstown v. Reeder}.\textsuperscript{127} In \textit{Reeder}, the town decided to establish a civic project, known as a Youth Court.\textsuperscript{128} As part of this project, the town’s Youth Court Executive Board formed various sub-committees, including a Constitution Committee and Steering Committee.\textsuperscript{129} Michael Reeder voluntarily served on each of these committees.\textsuperscript{130} As part of his role on the Steering Committee, Reeder was tasked with preparing a manual for the Youth Court.\textsuperscript{131} Reeder drafted the manual “after receiving conceptual and practical input from many people,” including guidance from members of the different committees and Executive Board.\textsuperscript{132} The chairman of the Executive Board relayed feedback from the committees to Reeder and served as a sounding board as Reeder prepared the manual.\textsuperscript{133}

Shortly after the manual was complete, Reeder contracted with another town to help them create a Youth Court and asserted a copyright in the Clarkstown manual.\textsuperscript{134} Initially, Reeder granted Clarkstown a license, but after the Executive Committee requested Reeder to assign the copyright to the town, Reeder revoked the license and litigation ensued.\textsuperscript{135} At issue was whether Reeder was an employee of the town and, as a result, whether the manual was a work made for hire with the copyright belonging to the town.\textsuperscript{136}

The court held that Reeder, despite the fact that he volunteered his time, was the town’s employee.\textsuperscript{137} The court held that the crucial factor in determining whether Reeder was an employee “is whether

\begin{itemize}
  \item \textsuperscript{124} Reid, 490 U.S. at 738.
  \item \textsuperscript{125} Town of Clarkstown v. Reeder, 566 F. Supp. 137, 141 (S.D.N.Y. 1983).
  \item \textsuperscript{126} See id. (citing Epoch Producing Corp. v. Killiam Shows, Inc., 522 F.2d 737, 744 (2d Cir. 1975); Roy Exp. Co. v. Columbia Broad. Sys., 503 F. Supp. 1137, 1149 (S.D.N.Y. 1980)).
  \item \textsuperscript{127} Id.
  \item \textsuperscript{128} Id. at 137-38.
  \item \textsuperscript{129} Id. at 138.
  \item \textsuperscript{130} Id.
  \item \textsuperscript{131} Id. at 139.
  \item \textsuperscript{132} Id.
  \item \textsuperscript{133} Id.
  \item \textsuperscript{134} Id.
  \item \textsuperscript{135} Id. at 139-40.
  \item \textsuperscript{136} Id. at 140-41.
  \item \textsuperscript{137} Id. at 141-42.
\end{itemize}
the alleged employer has the right to direct and supervise the manner in which the writer performs his work.” 138 The town had this right to direct and supervise the manner in which the manual was created because (1) the Chairman of the Executive Board had the power to remove and appoint members; (2) Reeder and the Chairman had ongoing discussions about drafting the manual, including what to include and change; (3) the Chairman assigned Reeder and another volunteer to prepare other documents and they did; and (4) Reeder submitted proposals to the committees and Executive Board for approval. 139 The fact that no one from the town ever ordered Reeder to write the manual in a specific format was not dispositive. 140 The court emphasized that the employment relationship depended on having the right to control, not exercising the right. 141

The court’s justification for this interpretation of the term “employee” in the 1976 Act leaves much to be desired. Rather than looking at the legislative history of the 1976 Act, the court reached its conclusion about how to distinguish employees from independent contractors by relying on three cases interpreting the muddled 1909 Act. 142

2. Actual Control Test

The second test used by the courts following enactment of the 1976 Act was the “actual control” test. 143 This was a variation of the right to control test. 144 Under the actual control test, courts did not just look at whether the hiring party could control or direct the work, but looked at whether the hiring party actually asserted control over the creation of the work. 145

The leading case on the actual control test is the Second Circuit’s Aldon Accessories Ltd. v. Spiegel, Inc. 146 In that case, the plaintiff, Aldon, was in the business of designing and selling figurines. 147 One of Aldon’s principals, Arthur Ginsberg, handled the creative aspects of the company, including product-design. 148 In anticipation of a new

---

138. Id. at 141 (internal quotations omitted).
139. Id.
140. Id. at 142.
141. Id.
142. Id. at 140-41; see supra final sentence in Part II.B.
144. Id. at 738-39.
145. Id. at 742; Landau, supra note 66, at 123.
146. 738 F.2d 548 (2d Cir. 1984).
147. Id. at 549.
148. Id.
line of mythological porcelain statuettes, Ginsberg contacted a Japanese firm, Wado, about making models for the porcelain statuettes.\textsuperscript{149}

In addition to describing the pose of the porcelain statuettes and sending drawings to Wado, Ginsberg traveled to Japan and worked with the Wado artists to develop the porcelain statuette models.\textsuperscript{150} Working with the artists included Ginsberg spending “hours and hours changing shapes, adjusting attitudes and proportions” in addition to giving specific directions to the artists on where to position the figures’ heads, legs, and hair.\textsuperscript{151} Afterwards, Aldon decided to develop brass versions of the statuettes and contacted a Taiwanese firm, Unibright, about creating brass models that differed from the porcelain models.\textsuperscript{152} Ginsberg traveled to Taiwan and worked with Unibright’s artists in the same way he worked with Wado’s artists.\textsuperscript{153}

After Aldon began selling the statuettes, a buyer for Spiegel inspected the statuettes at a trade show and requested a sample.\textsuperscript{154} Aldon sent the samples and a few months later discovered Spiegel selling identical statuettes in its catalogue.\textsuperscript{155} Aldon sued for copyright infringement and prevailed at trial, including on the issue of whether the models created by Wado and Unibright were works made for hire.\textsuperscript{156}

On appeal, Spiegel argued that Wado and Unibright, not Aldon, were the copyright owners (i.e. the models were not works made for hire).\textsuperscript{157} The Second Circuit disagreed and held that the Wado and Unibright artists were employees working within the scope of their employment, because “Ginsberg did much more than communicate a general concept or idea to the [artists], leaving creation of the expression solely to them.”\textsuperscript{158} The court was persuaded by the fact that “Ginsberg actively supervised and directed the creation” of the designs and that although he did not “physically wield the sketching pen and sculpting tools, he stood over the artists and artisans at critical stages of the process, telling them exactly what to do.”\textsuperscript{159}

In adopting the actual control test, the Second Circuit cited to several cases interpreting the 1909 Act and noted that there was “[n]othing in the 1976 Act or its legislative history indicat[ing] that

\begin{itemize}
\item \textsuperscript{149} Id.
\item \textsuperscript{150} Id.
\item \textsuperscript{151} Id. at 550.
\item \textsuperscript{152} Id.
\item \textsuperscript{153} Id.
\item \textsuperscript{154} Id.
\item \textsuperscript{155} Id.
\item \textsuperscript{156} Id. at 550-51.
\item \textsuperscript{157} Id. at 551.
\item \textsuperscript{158} Id. at 553.
\item \textsuperscript{159} Id.
\end{itemize}
Congress intended to dispense with this prior law . . . ”160 Rather than dissecting the legislative history as Professor Hardy has done, the court concluded that Congress did not intend to narrow what was meant by “employee,” because there would surely have been some discussion of this in the legislative history.161 In addition to the Second Circuit, the actual control test was adopted by the Fourth and Seventh Circuits.162

3. Agency Test

The third test used by the courts following enactment of the 1976 Act was the “agency” test.163 Under this test, courts determined whether a hired party was an employee or independent contractor by using the meaning of the word “employee” as understood under agency law.164 Although there was no federal agency law, the courts suggested using the Restatement (Second) of Agency as a guide.165 According to the Restatement, courts consider the following factors in distinguishing between employees and independent contractors:

(a) the extent of control which, by the agreement, the master may exercise over the details of the work;
(b) whether or not the one employed is engaged in a distinct occupation or business;
(c) the kind of occupation, with reference to whether, in the locality, the work is usually done under the direction of the employer or by a specialist without supervision;
(d) the skill required in the particular occupation;
(e) whether the employer or the workman supplies the instrumentalities, tools, and the place of work for the person doing the work;
(f) the length of time for which the person is employed;
(g) the method of payment, whether by the time or by the job;

160. Id. at 552.
161. Id.
162. Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 739 (1989); see Brunswick Beacon, Inc. v. Schock-Hopchas Publ’g Co., 810 F.2d 410, 413 (4th Cir. 1987) (“[T]here is no suggestion that they supervised Beacon employees as they developed the advertisements or directed the manner of the work’s completion.”); Evans Newton, Inc. v. Chicago Sys. Software, 793 F.2d 889, 894 (7th Cir. 1986) (“We find the Second Circuit’s analysis and conclusion [in Aldon] compelling.”). But see Landau, supra note 66, at 127 (arguing that although the Seventh Circuit adopted the actual control test from Aldon, it applied it in a way that more closely conforms to the right to control test).
163. Reid, 490 U.S. at 739.
164. Easter Seal Soc’y for Crippled Children & Adults of La., Inc. v. Playboy Enter., 815 F.2d 323, 334-35 (5th Cir. 1987).
165. Id. at 335.
(h) whether or not the work is part of the regular business of the employer;

(i) whether or not the parties believe they are creating the relation of master and servant; and

(j) whether the principal is or is not in business.\footnote{166}

The leading case using the agency test was the Fifth Circuit’s decision in \textit{Easter Seal}.\footnote{167} \textit{Easter Seal} involved a videotape of a staged Mardi Gras style parade.\footnote{168} A representative of the Easter Seal Society entered into a contract with a television station to film the parade and edit it to be a sixteen-minute segment for the Easter Seal Society’s telethon.\footnote{169} The Easter Seal Society representative gave some suggestions to Mr. Beyer, the head of the film crew, including camera locations, scenes to look for during the parade, and camera angles.\footnote{170} However, it was not clear whether Mr. Beyer followed these suggestions.\footnote{171} In addition, the Easter Seal Society did not control technical issues, such as lighting, sound, and color balance.\footnote{172} Mr. Beyer made all decisions on aesthetics and technical issues concerning the cameras and sound equipment.\footnote{173}

After being aired nationally, the television station permitted a Canadian television producer to use pieces of the film footage.\footnote{174} As it turned out, the Canadian producer used the film in an adult film, \textit{Candy, the Stripper}.\footnote{175} \textit{Candy, the Stripper} was distributed and shown nationally by Playboy and others.\footnote{176} After learning about this, Easter Seal Society filed suit alleging copyright infringement.\footnote{177} The district court held Mr. Beyer was not an employee of the Easter Seal Society, and therefore, the television station, which did employ Mr. Beyer, owned the copyright in the film.\footnote{178}

On appeal, the Fifth Circuit considered three different interpretations of what constituted an employee for work made for hire purposes and analyzed the problems associated with each.\footnote{179} The Fifth Cir-
cuit discussed a variation of the right to control test and described it as ignoring the distinction between employees and independent contractors.\textsuperscript{180} Under this test, the courts focus on whether the work was done at the instance and expense of the hiring party and ask whether the hiring party had the right to control the work.\textsuperscript{181} The Fifth Circuit criticized this test as making the nine categories of works under § 101(2) “completely mysterious”\textsuperscript{182} and also not deviating from the interpretation of work made for hire under the 1909 Act, despite a belief that Congress was trying to “tighten up the ‘work for hire’ doctrine under the 1976 Act.”\textsuperscript{183}

Likewise, the Fifth Circuit was critical of the Second Circuit’s actual control test in \textit{Aldon}.\textsuperscript{184} The Fifth Circuit was concerned that the actual control test could result in a lack of consistency between the same buyer and seller if more than one work were produced.\textsuperscript{185} Moreover, the Fifth Circuit thought the test was overly complicated, because if a formal, salaried employee were hired, but the work was not actually supervised by the employer, then the employer’s copyright would be lost and this “would be almost unimaginable.”\textsuperscript{186} Finally, the Fifth Circuit criticized the actual control test, because it easily slid into the right to control test.\textsuperscript{187}

The Fifth Circuit was also critical of the agency test.\textsuperscript{188} The court thought that such an interpretation was radically different from the work made for hire doctrine under the 1909 Act, and the court was not sure Congress had this in mind.\textsuperscript{189} Despite these concerns, the Fifth Circuit adopted the agency test because it: (1) made sense out of the specially commissioned works categories in § 101(2); (2) tied the meaning of “work made for hire” to a well-developed doctrine in agency law; (3) enhanced predictability; and (4) created a “moral symmetry” with other areas of the law, such as the employer being liable under the theory of respondeat superior.\textsuperscript{190}

\textsuperscript{180} \textit{Id.} at 331.
\textsuperscript{181} \textit{Id.} Note that the Fifth Circuit’s statement of the right to control test focuses on the right to control the work as opposed to the manner in which the work is performed, which was set forth in \textit{Reeder}. See supra Part II.D.1.
\textsuperscript{182} \textit{Easter Seal}, 815 F.2d at 331.
\textsuperscript{183} \textit{Id.}
\textsuperscript{184} \textit{Id.} at 333-34.
\textsuperscript{185} \textit{Id.} at 333.
\textsuperscript{186} \textit{Id.} at 334.
\textsuperscript{187} \textit{Id.}
\textsuperscript{188} \textit{Id.} at 330-31.
\textsuperscript{189} \textit{Id.}
\textsuperscript{190} \textit{Id.} at 335. Although the other justifications are probably self-evident, with respect to the fourth justification—moral symmetry—the Fifth Circuit explained “a buyer is a statutory ‘author’ if and only if he is responsible for the negligent acts of the seller. For
4. Formal, Salaried Employee Test

The final test used by the courts following enactment of the 1976 Act was the “formal, salaried employee” test.\textsuperscript{191} Under this test, courts initially looked to see if the hired party “[held] himself or herself out as a freelancer.”\textsuperscript{192} If so, then the hiring party should have anticipated the work not being a work made for hire.\textsuperscript{193} If the relationship was ambiguous, then a variety of factors were examined, most of which were a subset of the factors identified under the agency test.\textsuperscript{194} These factors included:

\begin{itemize}
  \item (1) whether the [hired party] worked in his or her own studio or on the premises of the [hiring party];
  \item (2) whether the [hiring party] is in the regular business of creating works of the type purchased;
  \item (3) whether the [hired party] works for several [hiring parties] at a time, or exclusively for one;
  \item (4) whether the [hiring party] retains authority to assign additional projects to the [hired party];
  \item (5) the tax treatment of the relationship by the parties;
  \item (6) whether the [hired party] is hired through the channels the [hiring party] customarily uses for hiring new employees;
  \item (7) whether the [hired party] is paid a salary or wages, or is paid a flat fee; and
  \item (8) whether the [hired party] obtains from the [hiring party] all benefits customarily extended to its regular employees.\textsuperscript{195}
\end{itemize}

Importantly, the formal, salaried employee test does not inquire into the degree of control and input the hiring party exercises.\textsuperscript{196}

The leading case using the formal, salaried employee test was the Ninth Circuit’s decision in \textit{Dumas v. Gommerman}.\textsuperscript{197} In \textit{Dumas}, ITT Cannon contracted with graphic artist Patrick Nagel to produce four paintings that ITT would give out as sets of lithographs to ITT Cannon’s distributors as part of a promotional campaign.\textsuperscript{198} ITT Cannon’s advertising agency determined the content and some parts of the design, borders, and placement of figures and gave Nagel sketches for him to use in making the illustrations.\textsuperscript{199} After paying for the paintings and distributing some lithograph sets, ITT Cannon had some lithographs leftover and eventually sold some to Stefan Gommerman,

---

\textsuperscript{191} Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 739 (1989).
\textsuperscript{192} \textit{Dumas v. Gommerman}, 865 F.2d 1093, 1105 (9th Cir. 1989).
\textsuperscript{193} \textit{Id.}
\textsuperscript{194} \textit{Id.}
\textsuperscript{195} \textit{Id.}
\textsuperscript{196} \textit{Id.}
\textsuperscript{197} \textit{Id.}
\textsuperscript{198} \textit{Id.} at 1094.
\textsuperscript{199} \textit{Id.}
an art gallery owner.\textsuperscript{200} In addition to purchasing the lithographs, Gommerman purchased ITT Cannon’s copyrights in the works.\textsuperscript{201}

Nagel’s widow, Jennifer Dumas, became the successor in interest in Nagel’s copyrights.\textsuperscript{202} After learning about ITT Cannon’s purported transfer of the copyrights to Gommerman, Dumas objected to Gommerman’s reproduction and sales of the paintings, claimed she was the copyright owner because Nagel was an independent contractor, and filed suit for copyright infringement.\textsuperscript{203} Gommerman defended by arguing that the paintings were works made for hire, because Nagel was ITT Cannon’s employee.\textsuperscript{204} The district court granted a preliminary injunction in favor of Dumas and Gommerman appealed to the Ninth Circuit.\textsuperscript{205}

On appeal, the Ninth Circuit perused the legislative history of the 1976 Copyright Act and acknowledged that the final work made for hire language in the statute was the result of a negotiated compromise.\textsuperscript{206} The court was reluctant to upset that compromise.\textsuperscript{207} In analyzing the legislative history, the Ninth Circuit determined that the negotiating parties used the term “employee” when referring to “a salaried worker in a long-term position.”\textsuperscript{208} Because ITT Cannon conceded that Nagel was not a formal, salaried employee, the paintings were not works made for hire, and the court affirmed the district court.\textsuperscript{209}

The Ninth Circuit recognized that other courts had utilized different tests but criticized those interpretations.\textsuperscript{210} In particular, the Ninth Circuit opined that the Second Circuit’s actual control test from \textit{Aldon} failed to recognize that the 1976 Act was trying to substantively change copyright law under the 1909 Act and that it distorted the balance struck in the negotiations between stakeholders.\textsuperscript{211} Furthermore, the Ninth Circuit joined the Fifth Circuit’s criticisms of the actual control and right to control tests.\textsuperscript{212} And although the Ninth Circuit largely agreed with the Fifth Circuit’s agency interpretation in \textit{Easter Seal}, it did object to the agency test as importing the actual control or right to control tests and making it difficult to de-

\begin{itemize}
\item \textsuperscript{200} Id.
\item \textsuperscript{201} Id.
\item \textsuperscript{202} Id.
\item \textsuperscript{203} Id. at 1094-95.
\item \textsuperscript{204} Id. at 1094.
\item \textsuperscript{205} Id. at 1095.
\item \textsuperscript{206} Id. at 1098-1101.
\item \textsuperscript{207} Id. at 1099.
\item \textsuperscript{208} Id. at 1101 (emphasis omitted).
\item \textsuperscript{209} Id. at 1102.
\item \textsuperscript{210} Id.
\item \textsuperscript{211} Id. at 1103.
\item \textsuperscript{212} Id.
\end{itemize}
termine, ex ante, whether the hired party was an employee or independent contractor. Instead, by conceiving of employees under the formal, salaried employee test, the Ninth Circuit believed there would be few disputes concerning the status of the hired party.

E. CCNV v. Reid and the Multifactor Balancing Test

Given the four different approaches to determining whether a hired party was an employee or independent contractor, it was no surprise the Supreme Court granted certiorari to resolve this circuit split. On June 5, 1989, the Supreme Court handed down its unanimous opinion in Community for Creative Non-Violence v. Reid and interpreted what Congress meant by the term “employee” in the work made for hire doctrine under the 1976 Act.

CCNV was a nonprofit organization committed to eliminating homelessness. In 1985, CCNV hired James Earl Reid to produce a sculpture for use in a Washington D.C. Christmas pageant. Members of CCNV conceived the idea of the sculpture as depicting a life-size nativity scene, but instead of using the Holy Family, the family would be homeless people huddled on a steam grate. CCNV and Reid never signed a written agreement nor did they discuss copyright ownership.

During the course of creating the sculpture, CCNV and Reid communicated several times about the sculpture design, including the position of the family, the items used to hold the family’s personal belongings, and who would serve as models for the family members. Reid eventually completed the sculpture and it was displayed near the pageant site for a month. The sculpture was returned to Reid for minor repairs while CCNV prepared to take the sculpture on tour. Reid objected to taking the sculpture on tour and when CCNV requested return of the sculpture, Reid refused. Both parties claimed ownership of the copyright, and CCNV filed suit seeking a declaration of ownership.

213. Id. at 1104.
214. Id. at 1105.
216. Id. at 733.
217. Id.
218. Id.
219. Id. at 734.
220. Id.
221. Id. at 735.
222. Id.
223. Id.
224. Id.
The district court, after a two-day bench trial, declared that the sculpture was a work made for hire. In particular, the district court held that Reid was an employee, because CCNV was the motivating factor in the production of the work and because CCNV directed enough of Reid’s efforts in creating the sculpture CCNV desired. On appeal, the D.C. Circuit reversed and held the sculpture was not a work made for hire. The D.C. Circuit adopted the Fifth Circuit’s agency test from Easter Seal and determined that Reid was an independent contractor rather than a CCNV employee.

The Supreme Court, in explaining the meaning of the work made for hire language, noted that the 1976 Act does not provide a definition of the term “employee” and that because of this ambiguity, the four interpretations discussed earlier had emerged. Ultimately, the Court adopted the agency test, because it was “well established that ‘[w]here Congress uses terms that have accumulated settled meaning under . . . the common law, a court must infer, unless the statute otherwise dictates, that Congress means to incorporate the established meaning of these terms.’” Because, the Court reasoned, Congress used the term “employee” in conjunction with the term “scope of employment”—a widely used term of art in agency law—and did not use any other language in the statute to indicate that it meant something other than the common law notion of the relationship between employers and employees, the common law agency test was appropriate.

The Court then fleshed out what it meant by the agency test. First, it stated that in determining whether a hired party is an employee, “we consider the hiring party’s right to control the manner and means by which the product is accomplished.” Next, citing section 220(2) of the Restatement (Second) of Agency, the Court listed the following factors as relevant to this inquiry:

1. the skill required; 2. the source of the instrumentalities and tools; 3. the location of the work; 4. the duration of the relationship between the parties; 5. whether the hiring party has the

225. Id.
226. Id. at 735-36.
227. Id. at 736.
228. Id. The D.C. Circuit also suggested the possibility that the sculpture could have been jointly authored by Reid and CCNV in which case both Reid and CCNV would own the copyright; the court remanded the case to the district court on this basis. Id.; see also 17 U.S.C. § 201(a) (2012) (describing authors of a joint work as co-owners of the copyright).
230. Id. at 739 (quoting NLRB v. Amax Coal Co., 453 U.S. 322, 329 (1981)).
231. Id. at 740.
232. Id. at 751-52.
233. Id. at 751.
right to assign additional projects to the hired party; [6] the extent of the hired party's discretion over when and how long to work; [7] the method of payment; [8] the hired party's role in hiring and paying assistants; [9] whether the work is part of the regular business of the hiring party; [10] whether the hiring party is in business; [11] the provision of employee benefits; and [12] the tax treatment of the hired party. 234

Although the Court cited the Restatement for the test, it is interesting to note that the factors listed in Reid do not match up exactly to those factors in the Restatement. 235 Besides these oversights, the Court failed to provide any guidance as to how these factors should be balanced other than noting that “[n]o one of these factors is determinative.” 236 The Court then applied these factors to the facts of the case and held that Reid was an independent contractor, not an employee, of CCNV, because although CCNV directed Reid’s work to the extent it met their specifications, all of the other factors weighed in favor of Reid being an independent contractor. 237

In the course of justifying the common law agency test, the Court explained that this interpretation furthered “Congress’ paramount goal in revising the 1976 Act of enhancing predictability and certainty of copyright ownership.” 238 In contrast, the actual control test undermined predictability and certainty, because the parties would not be able to know until late in the process whether the hiring party had actually wielded sufficient control over the hired party. 239 Because CCNV had conceded that the closely-related right to control test was difficult to demonstrate without actual control, the right to control test suffered from a similar flaw. 240

In addition to furthering the policy goals of predictability and certainty, the Court explained that the right to control test focused on

---

234. Id. at 751-52 (citing RESTATEMENT (SECOND) OF AGENCY § 220(2) (1958)).
235. Compare id., with RESTATEMENT (SECOND) OF AGENCY § 220(2) (1958); see also Assaf Jacob, Tort Made For Hire – Reconsidering the CCNV Case, 11 YALE J.L. & TECH. 96, 109 (2009). The Court failed to include some of the Restatement factors and added new factors not listed in the Restatement. Id. For example, the Court introduced the following factors: “the right to assign additional projects to the hired party; the extent of the hired party’s discretion over when and how long to work; the hired party’s role in hiring and paying assistants; and the provision of employee benefits and tax treatment of the hired party.” Id. Likewise, “the Court omitted the following factors listed in the Restatement: whether or not the hired party is engaged in a distinct occupation or business; the kind of occupation, with reference to whether, in the locality, the work is usually done under the direction of the employer or by a specialist without supervision; and whether or not the parties believe they are creating the relations of master and servant.” Id. at 110.
236. Reid, 490 U.S. at 752.
237. Id. at 752-53.
238. Id. at 749.
239. Id. at 750.
240. Id. at 750 n.17.
the relationship between the hiring party and the product, rather than the relationship between the hiring and hired parties. Accordingly, this focus was misguided, because the work made for hire language in section 101 is written in terms of the latter rather than former. Moreover, the right to control test ignored the dichotomy between works created by employees and specially commissioned works, because under the right to control test, a specially commissioned work could also be a work by an employee as long as the hiring party had the right to control the product. In short, the right to control test would largely eliminate the statutory requirement for specially commissioned works of having a signed writing specifying the product is a work made for hire.

Notwithstanding the Court’s critiques of the right to control test, the test set forth in Reid seemed to adopt that test as the ultimate question in the analysis. The confusion in the Court rejecting the right to control test while at the same time partially adopting it is that the Court explained the right to control test differently than the way it was originally stated in Reeder. Under Reeder, the court defined the test as “whether the alleged employer has the right to direct and supervise the manner in which the [hired party] performs his work.” This test is nearly identical to the test the Supreme Court partially adopts in Reid. The right to control test rejected in Reid was slightly different. The Supreme Court’s criticisms were aimed at a test focused on whether “the hiring party retain[ed] the right to control the product.” The difference between these two versions of the right to control test is that one focuses on the hiring party controlling the manner and means of production, whereas the other focuses on the hiring party controlling the product. The cases the Supreme Court cites to in its discussion of the right to control test framed the test in terms of the hiring party controlling the manner and means of production, rather than the product itself, but whether this is a meaningful distinction is doubtful.
The actual control test suffered from a similar problem of failing to dichotomize. The Court noted that although a work could be a work made for hire under section 101(2) and not under 101(1) if the work was specially commissioned, but no actual control was exercised, the Court found there was no support for this distinction in the statutory language.\(^{251}\) Finally, the Court summarily rejected the formal, salaried employee test, because although there was some support for this approach in the legislative history, the work made for hire provision used the term “employee” or “salaried employee.”\(^{252}\) In addition, the amici arguing for this approach did not agree on the standard for what constituted a “formal, salaried employee.”\(^{253}\)

In sum, the Court’s decision in *Reid* settled the question that had plagued the lower courts for several years. And, although the lower courts now had a multifactor test to help them determine whether the hired party was an employee or independent contractor, all they were told about the test was that no single factor was dispositive. This lack of further guidance led to additional confusion about how the factors ought to be applied.

**F. Weighting Factors—The Aymes v. Bonelli Three-Year Study**

Following the Supreme Court’s decision in *Reid*, the lower courts were confronted with how the *Reid* factors were to be applied. According to the Second Circuit, one district court erred in treating all the *Reid* factors as equally important and simply tallying the factors in making its determination.\(^{254}\) To provide the lower courts additional guidance in applying the *Reid* factors, the Second Circuit in *Aymes v. Bonelli* suggested that “there are some factors that will be significant in virtually every situation.”\(^{255}\) In particular, the *Aymes* court listed the following *Reid* factors as almost always being relevant and deserving of more weight in the multifactor analysis:

1. the hiring party’s right to control the manner and means of creation;\(^{256}\)
2. the skill required;
3. the provision of employee

\(^{251}\) *Reid*, 490 U.S. at 742 (“[T]here is no statutory support for an additional dichotomy between commissioned works that are actually controlled and supervised by the hiring party and those that are not.”).

\(^{252}\) *Id.* at 742 n.8. It is worth noting that section 101 also does not use the term employee as understood by the common law of agency.

\(^{253}\) *Id.*


\(^{255}\) *Id.*

\(^{256}\) Read literally, the hiring party’s right to control the manner and means of creation is not one of the *Reid* factors, but it is the ultimate question that the *Reid* factors help courts determine. As the Supreme Court stated in *Reid*, “[i]n determining whether a hired party is an employee under the general common law of agency, we consider the hiring party’s right to control the manner and means by which the product is accomplished. Among
benefits; (4) the tax treatment of the hired party; and (5) whether the hiring party has the right to assign additional projects to the hired party.\textsuperscript{257}

Of these important factors, the Second Circuit additionally noted that the employee benefits and tax treatment factors were especially probative, because “every case since Reid that...applied the test...found the hired party to be an independent contractor where the hiring party failed to extend benefits or pay social security taxes.”\textsuperscript{258} In determining which Reid factors were deserving of more weight, the Aymes court did not simply pull these five factors out of thin air. Instead it relied on all of the cases decided in the intervening three and a half years since Reid—six cases in total.\textsuperscript{259}

With respect to the remaining Reid factors, the Aymes court indicated that some were generally of little use in the work made for hire analysis.\textsuperscript{260} In particular, the court noted that whether the work is part of the regular business of the hiring party and whether the hiring party is in business will generally have little weight in the analysis.\textsuperscript{261}

III. UPDATING AYMES—A COMPREHENSIVE MULTIFACTOR STUDY

The pronouncement of the important factors in Aymes took place a mere three years after Reid. Since Aymes, many more work made for hire cases have been decided. Accordingly, a fresh and more comprehensive look at courts’ analyses of the Reid factors will facilitate a greater understanding of which factors are the most important in determining whether a hired party is an employee or independent contractor. This study examines the universe of work made for hire cases since Reid where a determination of the hired party’s status has been made and evaluates which factors are and are not important in the analysis. This part sets forth the methodology for locating and coding the cases. Next, it describes the results of this study and clusters the factors based on their importance.

\textsuperscript{257} Aymes, 980 F.2d at 861.

\textsuperscript{258} Id. at 863.

\textsuperscript{259} Id. at 860-63.

\textsuperscript{260} Id. at 863-64.

\textsuperscript{261} Id. at 863 (“[T]his factor will generally be of little help in this analysis.”).
A. Methodology

Using a broad search, I located what I believe to be all of the work made for hire cases decided under the 1976 Copyright Act from June 5, 1989 through June 5, 2014. After eliminating false positives, this search yielded forty-six work made for hire cases. For each case, the following information was recorded: whether the court concluded the hired party was an employee or independent contractor, the identity of the court and which circuit it fell within, the year the case was decided, and whether the case cited Aymes in the context of the five factors the Aymes court found to almost always be important.

In addition, for each case two variables were created for each Reid factor. The first variable was whether the factor favored employee status, favored independent contractor status, was indeterminate, or was not addressed by the court. These determinations were made based on statements by the courts about how these factors affected the outcome.

The second variable was whether the court weighted each Reid factor. The coding options for this variable were that the court gave

---

262. I ran the following search in the ALLCASES database in Westlaw, which contains all federal and state court cases: employee /p ("work for hire" "work made for hire") /p factor!. I then limited the search results to the date range between June 5, 1989 and June 5, 2014. This yielded 109 cases. Of these 109 cases, 43 analyzed the Reid factors and determined whether the hired party was an employee or independent contractor. I then ran this search in Lexis: employee /para ("work for hire" or "work made for hire") /para factor!. I then limited the search results to the date range between June 5, 1989 and June 5, 2014. The Lexis search produced an additional eighteen cases that were not already captured in the Westlaw search. Only two of these eighteen cases produced a copyright case that analyzed the Reid factors and determined whether the hired party was an employee or independent contractor. I then ran this search in Bloomberg Law: employee p/ ("work made for hire" OR "work for hire") p/ factor!. I then limited the search results to the date range between June 5, 1989 and June 5, 2014. The Bloomberg search produced an additional five cases not already captured in the Westlaw and Lexis searches. Only one of these five cases produced copyright cases that analyzed the Reid factors and determined whether the hired parties were employees or independent contractors. I recognize that there might be other work made for hire cases that have been decided, but were not reported in Westlaw, Lexis, or Bloomberg. See generally Elizabeth Y. McCuskey, Submerged Precedent, passim (unpublished manuscript) (on file with author) (describing the concept of submerged precedent); Elizabeth Y. McCuskey, Clarity and Clarification: Grable Federal Questions in the Eyes of Their Beholders, 91 Neb. L. Rev. 387, 427-30 (2012) (same). Given the limited search capabilities of docket management systems, it is cost-prohibitive to do any more of an exhaustive search.

263. The false positives were mostly non-copyright cases (e.g. Title VII) or copyright cases where the court did not decide whether the hired party was an employee or independent contractor (e.g. denying a motion for summary judgment because there was a genuine issue of material fact regarding the analysis).

264. In one case, Huebbe v. Oklahoma Casting Co., 663 F. Supp. 2d 1196 (W.D. Okla. 2009), the District Court engaged in two separate work made for hire analyses. Because these analyses were distinct, I treated them as two cases. This explains why although there were forty-five cases in the search, I analyzed forty-six cases as part of this study.

265. Aymes itself was also included in this last variable.
additional weight to the particular factor, discounted the factor, or
did not expressly weight the factor.\footnote{266}{The weighting or
discounting of the factors could be done either by the court stating
that the factor generally was entitled to additional weight or should
generally be discounted (i.e. the court stating that in its circuit, the
following factors are important) or during the application of the factor
to the facts of the particular case (i.e. the court not making a
pronouncement about the factors' importance in the abstract, but
stating that a particular factor was particularly important during the
analysis of the facts before it).} These too were based on state-
ments (or the absence of statements) by the courts about how much
weight they were giving each of the factors.

In addition to the twelve Reid factors, two factors that were
addressed with some frequency were added: (1) the hiring party’s
right to control the manner and means by which the product is
accomplished,\footnote{267}{See supra note 256.} and (2) how the hired
party and hiring party referred to the hired party. Finally, because the Reid factors are not an exhaust-
ive list, any additional factors the courts discussed in their analyses
were recorded. All of the additional factors were coded the same as
the two variables used for the Reid factors.\footnote{268}{My research assistant and I independently coded all of the variables and
then met in person to resolve any conflicts in the coding by reviewing the case.}

\section*{B. Results}

Using this data, the following calculations were made: how fre-
quently each factor was addressed by the courts, how consistent each
factor was with the ultimate result about the status of the hired party,
and how frequently each factor was given additional weight or dis-
counted in the courts’ analyses. These calculations are shown below in Table 1 as Frequency,\footnote{269}{Frequency for each factor is calculated as forty-six total cases minus the number of cases the factor was not addressed by the courts.} Consistency,\footnote{270}{Consistency for each factor is calculated by adding together the total number of cases where the factor's outcome is consistent with the court's ultimate conclusion about the hired party's status. For example, if the skill required factor favors a finding of employee status and the court ultimately concludes the hired party is an employee, then this is consistent. If, however, the court finds that the skill required factor favors a finding of employee status, but the court ultimately concludes the hired party is an independent contractor, then this is inconsistent. The percentage in parentheses is calculated by dividing this number by the number in the Frequency column. In other words, when the factor is addressed, how consistent is it with the ultimate conclusion?} Favored Weighting,\footnote{271}{Favored Weighting for each factor is calculated by adding together the total number of cases where the factor is given additional weight. The percentage in parentheses is calculated by dividing this number by the number in the Frequency column. In other words, when this factor is addressed, how often is it favored?} and Discounted Weighting,\footnote{272}{Discounted Weighting for each factor is calculated the same way Favored Weighting is, but counts cases where the factor is discounted rather than given additional weight.} respectively.
These calculations all measure, in some form, the importance of the factors. The frequency with which a factor is analyzed suggests whether courts are considering the factor in the first place or simply ignoring it. A frequently ignored factor is less likely to be important than one frequently addressed. Likewise, consistency illustrates importance, because a factor that is less reliable in predicting the ultimate outcome suggests that courts treat that factor as having less impact on the ultimate result than other factors. Finally, whether a factor is discounted or given additional weight in the analysis reflects its importance as the courts are directly addressing which factors they take more seriously and find more probative in the analysis and which factors they routinely declare to be of less importance.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Frequency</th>
<th>Consistency</th>
<th>Favored Weighting</th>
<th>Discounted Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill required</td>
<td>26 (57%)</td>
<td>22 (85%)</td>
<td>8 (31%)</td>
<td>1 (4%)</td>
</tr>
<tr>
<td>Source of instrumentalities and tools</td>
<td>33 (72%)</td>
<td>29 (88%)</td>
<td>2 (6%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>Work location</td>
<td>32 (70%)</td>
<td>24 (75%)</td>
<td>0 (0%)</td>
<td>4 (13%)</td>
</tr>
<tr>
<td>Relationship duration</td>
<td>29 (63%)</td>
<td>23 (79%)</td>
<td>1 (3%)</td>
<td>2 (7%)</td>
</tr>
<tr>
<td>Additional projects</td>
<td>29 (63%)</td>
<td>25 (86%)</td>
<td>6 (20%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>When and how long to work</td>
<td>32 (70%)</td>
<td>26 (81%)</td>
<td>0 (0%)</td>
<td>2 (6%)</td>
</tr>
<tr>
<td>Payment method</td>
<td>39 (85%)</td>
<td>33 (85%)</td>
<td>5 (13%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Hiring and paying assistants</td>
<td>18 (39%)</td>
<td>13 (72%)</td>
<td>0 (0%)</td>
<td>2 (11%)</td>
</tr>
<tr>
<td>Part of regular business of hiring party</td>
<td>29 (63%)</td>
<td>24 (83%)</td>
<td>0 (0%)</td>
<td>4 (14%)</td>
</tr>
<tr>
<td>Hiring party in business</td>
<td>9 (20%)</td>
<td>4 (44%)</td>
<td>0 (0%)</td>
<td>3 (33%)</td>
</tr>
<tr>
<td>Employee benefits</td>
<td>36 (78%)</td>
<td>32 (89%)</td>
<td>12 (33%)</td>
<td>2 (6%)</td>
</tr>
<tr>
<td>Tax treatment</td>
<td>37 (80%)</td>
<td>32 (86%)</td>
<td>12 (32%)</td>
<td>2 (5%)</td>
</tr>
<tr>
<td>Right to control manner and means</td>
<td>36 (78%)</td>
<td>21 (58%)</td>
<td>5 (14%)</td>
<td>4 (11%)</td>
</tr>
<tr>
<td>Label</td>
<td>9 (20%)</td>
<td>5 (56%)</td>
<td>1 (11%)</td>
<td>3 (33%)</td>
</tr>
</tbody>
</table>

Tables 2, 3, 4, and 5 show how each factor fares with respect to each of the four calculations. As illustrated in those tables, some factors tend to rise toward the top of the list regardless of which calculations are being used. Inconsistent factors are also less useful to the parties and attorneys in predicting outcomes.

273. Inconsistent factors are also less useful to the parties and attorneys in predicting outcomes.
tion is used. Other factors are stable in the middle or at the bottom of the list. But there is some fluctuation of the factors’ rankings across calculations.

<table>
<thead>
<tr>
<th>Table 2 – Sorted by Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>Payment method</td>
</tr>
<tr>
<td>Tax treatment</td>
</tr>
<tr>
<td>Employee benefits</td>
</tr>
<tr>
<td>Right to control manner and means</td>
</tr>
<tr>
<td>Source of instrumentalities and tools</td>
</tr>
<tr>
<td>Work location</td>
</tr>
<tr>
<td>When and how long to work</td>
</tr>
<tr>
<td>Relationship duration</td>
</tr>
<tr>
<td>Additional projects</td>
</tr>
<tr>
<td>Part of regular business of hiring party</td>
</tr>
<tr>
<td>Skill required</td>
</tr>
<tr>
<td>Hiring and paying assistants</td>
</tr>
<tr>
<td>Hiring party in business</td>
</tr>
<tr>
<td>Label</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3 – Sorted by Consistency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>Employee benefits</td>
</tr>
<tr>
<td>Source of instrumentalities and tools</td>
</tr>
<tr>
<td>Additional projects</td>
</tr>
<tr>
<td>Tax treatment</td>
</tr>
<tr>
<td>Skill required</td>
</tr>
<tr>
<td>Payment method</td>
</tr>
<tr>
<td>Part of regular business of hiring party</td>
</tr>
<tr>
<td>When and how long to work</td>
</tr>
<tr>
<td>Relationship duration</td>
</tr>
<tr>
<td>Work location</td>
</tr>
<tr>
<td>Hiring and paying assistants</td>
</tr>
<tr>
<td>Right to control manner and means</td>
</tr>
<tr>
<td>Label</td>
</tr>
<tr>
<td>Hiring party in business</td>
</tr>
</tbody>
</table>

274. For example, payment method, employee benefits, and tax treatment are in the top half for each calculation.

275. For example, relationship duration is ranked eighth, ninth, ninth, and eighth in Tables 2, 3, 4, and 5, respectively.

276. For example, whether the hiring party is in business is ranked last in every calculation.

277. For example, payment method is ranked first when sorted by frequency and discounted weighting, but is only sixth for consistency and favored weighting.
With respect to weighting factors, it is worth noting that other than giving additional weight to employee benefits, tax treatment, skill required, and arguably the right to assign additional projects, courts do very little favorable weighting.\textsuperscript{278} Discounted weighting is even rarer. Although the hiring party being in business and the label used to describe the hired party are discounted thirty-three percent of the time, in terms of raw numbers, this discounting occurred in only three cases for each measure.\textsuperscript{279}

<table>
<thead>
<tr>
<th>Table 4 – Sorted by Favored Weighting (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Employee benefits</td>
</tr>
<tr>
<td>Tax treatment</td>
</tr>
<tr>
<td>Skill required</td>
</tr>
<tr>
<td>Additional projects</td>
</tr>
<tr>
<td>Right to control manner and means</td>
</tr>
<tr>
<td>Payment method</td>
</tr>
<tr>
<td>Label</td>
</tr>
<tr>
<td>Source of instrumentalities and tools</td>
</tr>
<tr>
<td>Relationship duration</td>
</tr>
<tr>
<td>Work location</td>
</tr>
<tr>
<td>When and how long to work</td>
</tr>
<tr>
<td>Hiring and paying assistants</td>
</tr>
<tr>
<td>Part of regular business of hiring party</td>
</tr>
<tr>
<td>Hiring party in business</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5 – Sorted by Discounted Weighting (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Payment method</td>
</tr>
<tr>
<td>Source of instrumentalities and tools</td>
</tr>
<tr>
<td>Additional projects</td>
</tr>
<tr>
<td>Skill required</td>
</tr>
<tr>
<td>Tax treatment</td>
</tr>
<tr>
<td>When and how long to work</td>
</tr>
<tr>
<td>Employee benefits</td>
</tr>
<tr>
<td>Relationship duration</td>
</tr>
<tr>
<td>Hiring and paying assistants</td>
</tr>
<tr>
<td>Right to control manner and means</td>
</tr>
<tr>
<td>Work location</td>
</tr>
<tr>
<td>Part of regular business of hiring party</td>
</tr>
<tr>
<td>Hiring party in business</td>
</tr>
<tr>
<td>Label</td>
</tr>
</tbody>
</table>

\textsuperscript{278} See Table 4.

\textsuperscript{279} See Table 5.
Although each calculation by itself is helpful in understanding the importance of a particular factor in the work made for hire analysis, no single calculation can tell the entire story. For example, if we only looked at the frequency with which the factors are addressed, then we might conclude that the skill required is one of the least important of the *Reid* factors. But although frequency tells us how often the courts address the factor, it could be that, when they do address it, they give it additional weight in the analysis, because the courts consider it to be important in the inquiry. The skill required factor falls into this category as it is weighted favorably 31% of the time. As described earlier, looking only at the weighting calculations is also of limited use because of the small number of cases where weight is discussed. As a result, all of the calculations must be examined together to discover which factors are the most and least important. Based on all four calculations in Table 1, I propose that the continuum shown in Figure 1 describes the relative importance of the work made for hire factors.

![Figure 1](image)

This continuum groups together factors that share similar features with respect to each of the calculations. For example, the factors listed as the most important (tax treatment, employee benefits, and payment method) all have high consistency (85%–89%), are addressed very frequently (78%–85%), are often or sometimes favorably weighted (13%–33%), and are never or infrequently discounted (0%–6%). The next group of factors (additional projects, skill required, and source of the instrumentalities and tools) all have a high consistency (85%–88%) but have a lower frequency (57%–72%); they do, however, have significant weighting (6%–31% favorable; 3%–4% discounting). Because this group is similar in many respects to the “most im-
important” group, but is addressed with less frequency, these factors are grouped together and deemed slightly less important.

At the other end of the continuum are the least important factors. These factors (right to control the manner and means, the label used to describe the hired party, and whether the hiring party is in business) have very poor consistency (44%-58%). And although the right to control the manner and means factor is addressed with high frequency (78%), this means very little given how inconsistent it is (58%) and that courts also split on whether it receives favorable weighting (14%) or is discounted (11%). The other two factors in this group are rarely addressed (20%) and are the most likely to be discounted in the analyses (33%).

Whether a particular factor belongs in one group down or one group up is certainly debatable; reasonable minds can differ. Illustrating the importance of the factors using a continuum with fuzzy lines separating the groups rather than strict lines of demarcation between the groups was purposefully chosen to acknowledge this.

To test how useful these groups are at predicting the ultimate outcome in a case, the top three factors were analyzed. If a majority of the three factors that were addressed favored the ultimate outcome, this was considered a successful prediction. If only two of the three factors were addressed and they were split, this was considered an unsuccessful prediction. Using only the three most important factors (tax treatment, employee benefits, and payment method), eighty-seven percent of the cases would be decided consistent with courts’ ultimate conclusion. When this study was done using the most important two groups, ninety-one percent of the cases would be decided consistent with the courts’ ultimate conclusion. In contrast, doing the same analysis using the three least important factors yielded a forty-eight percent success rate. Expanding this to the five least important factors improved the success rate to sixty-three percent. Doing the same analysis for the middle three factors yielded a seventy-eight percent success rate. These results are displayed below in Figure 2. In sum, the structure of this continuum appears to correctly illustrate the relative importance of the Reid and other commonly-considered factors.  

---

282. There are, of course, limitations on this study. The primary limitation is the small number of cases (n=46) involved in the study. Because of this relatively small number of cases, it makes it nearly impossible to make the data any more granular. For example, it might be interesting to see if the factors’ importance changes based on the type of industry or work at issue. Although there are some industries or works that occur with some frequency (e.g. architecture and software), there are not enough of these cases to be able to draw any meaningful conclusions. Another limitation of this study is the potential for selection bias. That is, the parties or their attorneys deciding which case to bring and not bring based on how the courts addressed the factors in previous decisions. This issue exists
IV. DISCUSSION

The results of this study lead to two broad questions. First, how do these results fit within the long and varied developments of the work made for hire doctrine? Second, going forward, what are the implications of these results? This part explores each of these questions.

With respect to reflecting on the results in light of the historical developments, this part discusses how these results square with the Second Circuit’s statement in Aymes about the importance and lack of importance of specific factors. In addition, it examines whether and how these results cohere with the pre-Reid interpretations of “employee” and the Supreme Court’s rationales for adopting and rejecting these interpretations.

With respect to the implications of these results, this part first explores how these results may affect general copyright litigation where the parties contest the work made for hire status of a work. Afterwards, these results are situated within the context of the impending litigation concerning the work made for hire status of music recorded after 1977. Finally, this part suggests how these results

with any study based on reported cases, but the effect could be more pronounced in this situation given the small number of cases.

283. See infra Part IV.A.
284. See infra Part IV.A.
285. See infra Part IV.B.1.
286. See infra Part IV.B.2.
may be of use to business planning attorneys so they can effectively advise their clients to achieve their desired results.  

A. Doctrinal Fit

In Aymes, the Second Circuit concluded that the following factors were the most important in the multifactor analysis:

(1) the hiring party’s right to control the manner and means of creation; (2) the skill required; (3) the provision of employee benefits; (4) the tax treatment of the hired party; and (5) whether the hiring party has the right to assign additional projects to the hired party.

Of the five Aymes factors, four are within the top two groups of the continuum. The only Aymes factor not in these top two groups is the hiring party’s right to control the manner and means of creation. This factor is surprisingly within the least important group of the continuum. As noted earlier, the Supreme Court in Reid suggested this was not itself a factor but was the ultimate question to be answered using the listed factors. Despite this, many lower courts have deemed the right to control a factor, rather than the ultimate conclusion. The lack of importance of this factor is surprising given that it is supposed to be determined by the other factors. Given that two courts (the Supreme Court and Second Circuit) have bolstered the importance of this factor, it is worth emphasizing that other courts have, by and large, ignored this authority.

In addition to overemphasizing the importance of the right to control factor, the Aymes court underappreciated the importance of the method of payment factor. The court acknowledged that this could be a “fairly important factor,” but as illustrated above, this factor turns out to be one of the three most important factors. To a lesser extent, the source of the instrumentalities was underappreciated by Aymes, which is within the same group as the skill required factor.

287. See infra Part IV.B.3.


289. Employee benefits and tax treatment are in the most important group. The skill required and right to assign additional projects are in the second most important group.

290. See supra note 256.

291. Id.

292. In theory, the lack of importance of the right to control factor could make sense given the way the Supreme Court and Restatement (Second) of Agency have phrased the test as the ultimate question. See Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 750-51 (1989); RESTATEMENT (SECOND) OF AGENCY §§220, 228 cmt. c (1958). In other words, if the lower courts interpreted Reid this way, then it would not be an important factor, because it is not really a factor at all. However, this explanation is not warranted given that the lower courts have frequently treated the right to control as a factor (seventy-eight percent frequency). See supra Table 1. Ultimately, this factor is not consistent with the outcomes and is treated inconsistently by the courts with respect to the weighting.

293. Aymes, 980 F.2d at 863.
In addition to declaring which factors were the most important, the Aymes court also opined that whether the work is part of the regular business of the hiring party and whether the hiring party is in business were of little use in the analysis. Based on the results of this study, it turns out that Aymes was certainly correct with respect to whether the hiring party is in business, because this factor is in the least important group. However, being part of the regular business of the hiring party is not as inconsequential as Aymes made it out to be. Although not one of the most important factors, this factor is in the middle of the pack.

In sum, the statements in Aymes about the importance or lack of importance of specific factors are largely, but not completely, consistent with the results of this study. Of course, this could be attributed to path dependency. That is, the courts were bound to follow or were heavily influenced by Aymes. Of the forty-six cases, thirty-six postdated Aymes. Of those thirty-six cases, eight (twenty-two percent) cite to Aymes, or a case citing Aymes, for the proposition that the Aymes factors are the important ones. Although path dependency probably had some influence on the ultimate outcome of the importance of the factors, the variations between the results of this study and the conclusions in Aymes suggest that path dependency was not dispositive in determining importance.

With respect to the results of this study as they relate to the pre-Reid tests for distinguishing employees from independent contractors, two observations are in order. First, as described earlier, although the Supreme Court cast the rejected right to control test as focusing on controlling the product as opposed to controlling the manner and means of production, it is questionable whether there is a meaningful difference between them. Nonetheless, the right to control test, as adopted by the lower courts before Reid, focused on the right to control the manner and means in which the hired party worked. Even assuming the Supreme Court had correctly understood the right to control test, the results of this study suggest that the right to control is unimportant. This factor is in the least important group on the continuum. As a result, the Supreme Court’s rejection of the right to control test is consistent with the results of this study. However, because the Supreme Court misunderstood the

294. Id.
295. Of course, courts that reached a similar conclusion with respect to how they balanced the factors could have read Aymes but failed to cite to the opinion.
296. See supra notes 246-49 and accompanying text.
297. See supra note 250 and accompanying text.
test, the reason for the consistency has little to do with the Court’s rationale for rejecting the test.299

Second, when comparing the formal, salaried employee test with the results of this study, we see that the Dumas factors tend to line up fairly well with the most important factors on the continuum. Of the eight Dumas factors, six are also found in Reid.300 Like Aymes, the Dumas factors account for four of the top six factors. But unlike Aymes, the Dumas factors account for all three of the most important factors. The other two Dumas factors that are found in Reid—whether the work is part of the regular business of the hiring party and location of the work—fall in the middle or lower end of the continuum, but not the lowest group as occurred with the Aymes factors. Moreover, one of the Dumas factors not listed in Reid—working for several hiring parties or exclusively for one—was addressed in four cases as an additional factor. This was the most common “other” factor and although not weighted favorably or discounted, it was 100% consistent with the ultimate outcomes in the cases. Given these results, it may be that courts are actually using a variation of the Dumas formal, salaried employee test. This comports with Professor Hardy’s view that the formal, salaried employee test is the appropriate one.301 Recall that Professor Hardy argued that the stakeholders assumed this was the test when they negotiated the terms of the 1976 Act in light of the 1909 Act, which seemed to recognize hiring party ownership of works created by salaried employees.302 The factors at the important end of the continuum support the view that a variation of the formal, salaried employee test is what the courts are actually using in arriving at their conclusions. This is an interesting result in light of how harshly the Supreme Court treated the formal, salaried employee test in Reid.303 The importance of these factors also

299. Interestingly, the Restatement (Third) of Agency, published in 2006, defines employee as “an agent whose principal controls or has the right to control the manner and means of the agent’s performance of work.” RESTATEMENT (THIRD) OF AGENCY § 7.07 (2006). The factors from the Restatement (Second) of Agency have been relegated to the comments of section 7.07. Id. at § 7.07 cmt. f. No court has suggested reviving the right to control test in light of this change in the Restatement.

300. These factors include: (1) whether the hired party worked in his or her own studio or on the premises of the hiring party; (2) whether the hiring party is in the regular business of creating works of the type purchased; (3) whether the hiring party retains authority to assign additional projects to the hired party; (4) the tax treatment of the relationship by the parties; (5) whether the hired party is paid a salary or wages, or is paid a flat fee; and (6) whether the hired party obtains from the hiring party all benefits customarily extended to its regular employees. Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 751-52 (1989).

301. See supra notes 99, 105, and accompanying text.

302. See supra notes 99, 105, and accompanying text.

303. See Reid, 490 U.S. at 742 n.8 (noting that the statutory definition of work made for hire “cannot support” the formal, salaried employee test, and acknowledging disagreement about the exact contours of the test). This test was rejected in a footnote, unlike the
brings us back full circle to the legislative history of the 1909 Act, which focused on employees being salaried.\textsuperscript{304}

One reason courts gravitate towards the three factors at the important end of the continuum could be that they are objectively measured and easy to apply. For example, whether the hiring party treated the hired party as an employee or independent contractor for tax purposes only involves looking at payroll documents and forms filed with the federal government. Likewise, providing benefits such as insurance is easily verified by looking at enrollment data. Similarly, the method of payment is easily determined by looking at paystubs to figure out the frequency and amounts paid. This is not to say that the other factors cannot be easily proven by looking at documents or other evidence, but factors such as skill required or the right to control the manner and means involve a certain level of subjectivity.

One final note about \textit{Reid} and the results of this study is apposite. In justifying the agency test, the Supreme Court explained that “Congress’ paramount goal in revising the 1976 Act [was] enhancing predictability and certainty of copyright ownership.”\textsuperscript{305} Given the results of this study and the multitude of cases where summary judgment is denied because of an issue of material fact about whether the hiring party is an employee or independent contractor, it is doubtful that the agency test has achieved this result.\textsuperscript{306} Under \textit{Reid}, courts balance about a dozen factors, some of which are consistent with the ultimate conclusion and some of which have very little to do with the ultimate result. It is odd to imagine a multifactor balancing test such as the one set forth in \textit{Reid} as providing more predictability and certainty to the parties than a smaller set of factors would achieve.\textsuperscript{307} If enhancing certainty and predictability of copyright ownership really were Congress’ paramount goals in the 1976 Act, then perhaps the

\textsuperscript{304} See supra notes 71-73 and accompanying text.

\textsuperscript{305} Reid, 490 U.S. at 749.

\textsuperscript{306} See, e.g., Int'l Code Council, Inc. v. Nat'l Fire Prot. Ass'n, No. 02C5610, 2006 WL 850879, at *23 (N.D. Ill. Mar. 27, 2006) (summary judgment is precluded because “a reasonable trier of facts could resolve the Reid multifactor analysis in either party's favor”); Brower v. Martin, 446 F. Supp. 2d 232, 235 (S.D.N.Y. 2006) (“In light of these and other disputed questions of material fact, the court cannot conclude as a matter of law whether plaintiff's songs constitute works-for-hire or to whom the copyright in these songs belongs.”); Ulloa v. Universal Music & Video Distrib. Corp., 303 F. Supp. 2d 409, 416 (S.D.N.Y. 2004) (“Having examined the Reid–Aymes factors and drawn all inferences in favor of the Plaintiff, the Court cannot conclude as a matter of law that either the musical composition or the sound recording of the Vocal Phrase were created as works for hire.”).

\textsuperscript{307} In the consumer context, we see that having too much choice or too many options can lead to bad results. See generally BARRY SCHWARTZ, THE PARADOX OF CHOICE – WHY MORE IS LESS passim (2004). Perhaps courts suffer from a similar version of analysis paralysis in that they either fail to resolve the issue themselves, and instead let juries decide, or latch onto a subset of factors and fail to engage in a complete multifactor analysis.
Supreme Court should reconsider the Reid test and adopt a subset of factors for the courts to consider. One option would be to only use the factors in the most important and second most important groups of the continuum. Another option would be to use the factors from the top two groups, but if those factors are evenly split, then turn to the less important factors as tie-breakers. In the event Congress has the opportunity to revisit the work made for hire doctrine and believes predictability and certainty are still valuable goals vis-à-vis copyright ownership, then perhaps Congress could provide more guidance as to what constitutes an employer-employee relationship by recognizing these factors as existing in tiers. Neither is likely, but given the lack of guidance from the Supreme Court on how to balance the factors, lower courts have the flexibility to adopt a formal recognition of the important and unimportant factors as suggested by the continuum presented here.

B. Practical Implications

Although the fit between the results of this study and the underlying doctrine and historical developments is intriguing, these results may also be of great value to litigators and business planning attorneys in the field. Where the parties dispute the work made for hire status of a copyright, these results will help litigators to gather and present evidence concerning the doctrine and to better evaluate their cases.308 One such instance, explored in more depth below, is the upcoming litigation between artists, producers, and record companies in regard to songs recorded after 1977.309 In the context of business planning, these results will help attorneys structure their clients’ relationships to help achieve the desired employment status.310 Each is discussed in detail below.

1. General Work Made for Hire Litigation

Litigation involving employment status under the work made for hire doctrine can arise in a few situations. For example, a defendant may assert that the plaintiff has no right to assert a claim for copyright infringement, because the plaintiff is not the legal or beneficial owner of the copyright as a result of the work being a work made for hire.311 Another example is a copyright infringement dispute between the hiring and hired parties where the alleged infringer successfully defends by claiming ownership of the copyright via the work made for hire.
hire doctrine. Another instance could be determining whether artists are entitled to assert violations of their moral rights under the Visual Artists Rights Act of 1990, which does not provide moral rights to works made for hire. Finally, the employment status in the work made for hire doctrine may arise in situations involving terminations of transfers, which also do not apply to works made for hire. In these situations, determining whether a party is an employee or independent contractor may entirely or partially dispose of a case or significantly shift negotiating power when trying to settle a dispute.

Armed with the results of this study, litigators engaged in discovery can focus their energy and clients’ money on the factors at the most important end of the continuum and pay less attention to those at the least important end. After completing discovery, the attorneys can use the continuum to decide how to present their arguments in favor of and against a work made for hire conclusion. For example, in a motion for summary judgment, rather than simply analyzing the factors in the order presented in , it may be more persuasive to analyze the factors in the order of importance. If most of the factors at the important end of the continuum support the movant, then presenting them this way may cause the judge to grant the motion. Likewise, if the factors at the more important end are split, the non-movant may want to present the factors like this to avoid summary judgment. In addition to gathering and presenting evidence of employment status, these results should assist attorneys in evaluating their clients’ cases and giving more informed recommendations about whether to settle a dispute and for how much.

312. See, e.g., Maness v. Heavrin, 97 F.3d 1457, 1457 (8th Cir. 1996) (unpublished table decision). The reason this situation arises is because one cannot infringe a copyright they own. Richmond v. Weiner, 353 F.2d 41, 42 (9th Cir. 1965) ("[A] copyright owner cannot infringe against his own copyright."").

313. 17 U.S.C. § 101(2)(B) (defining “work of visual art” as not including a work made for hire).

314. Id. § 203(a) (“In the case of any work other than a work made for hire, [certain granted rights are] subject to termination . . . .”). For a more detailed analysis on termination rights, see infra Part IV.B.2.

315. This is not to say that simply reordering the factors will cause a change in outcome in most cases. In fact, considering how to present the factors may only be useful in borderline cases where a court is on the fence about granting or denying a motion for summary judgment.

316. It is unclear if the parties would settle even if they knew this information. Looking at the 46 cases involved in this study, in 22 (48%) of them, 100% of the factors that were addressed and not indeterminate were consistent with the ultimate outcome in the case (e.g. if 9 factors were addressed, the court held all nine weighed in favor of the outcome). In 10 of the cases (22% of the total), only one factor was inconsistent with the ultimate conclusion. Combined, 70% of the cases involved a situation where 0 or 1 factor was inconsistent with the ultimate outcome. Once the attorneys in these cases knew what the facts were, it is hard to understand why the attorneys for the eventually-losing parties would push ahead with this aspect of the litigation.
the most important factors can provide the increased predictability and certainty that the Court in *Reid* thought it was providing.

2. **Specific Work Made for Hire Litigation—Terminations of Transfers and Sound Recordings**

One of the most interesting areas of copyright law where the work made for hire status will arise is with respect to terminations of transfers in sound recordings. Section 203 of the Copyright Act permits termination right owners to end certain post-1977 transfers of a copyright and reclaim the copyright for themselves. That is, if the work and type of transfer are eligible for termination and the termination right is exercised, all the rights covered by the transfer revert back to the owner of the termination interest. As a result, the termination right holder gets “a second bite at the apple” and can attempt to negotiate a new license or assignment on more favorable terms.

For example, imagine an unknown artist records a song at her home studio and in 2014 successfully sells her copyright in the sound recording to a record company for a measly sum. The recording turns out to be a huge success, and the record company earns millions of dollars from exploiting its acquired rights in the recording. Under section 203, after thirty-five years, the recording artist could reclaim the copyright in the sound recording. If the record company wanted to continue exploiting the recording, it would be forced to sit down at the negotiating table with the recording artist and reach a new deal. Presumably, the recording artist will negotiate a better deal in 2049 than she did in 2014.

As illustrated by this example, the termination of transfer right is a powerful one. As it turns out, 2013 was the first year of terminations under section 203 and a wave of termination disputes is expected over the next several years. These disputes (and the inev-

---

318. *Id.* § 203(b).
319. Corey Field, *Their Master’s Voice? Recording Artists, Bright Lines, and Bowie Bonds: The Debate Over Sound Recordings as Works Made For Hire*, 48 J. COPYRIGHT SOC’Y U.S.A. 145, 155 (2000). Sections 304(c) and (d) of the Copyright Act are additional termination of transfers provisions, but apply to copyrights transferred before 1978. 17 U.S.C. § 304(c), (d). Like section 203, these sections also exclude works made for hire from termination. But because the works were created before 1978, the 1909 Copyright Act’s work made for hire doctrine applies rather than the modern doctrine.
320. Section 203 applies to transfers made on or after January 1, 1978. § 203(a). The five-year termination window begins thirty-five years after execution of the transfer. Therefore, the earliest five-year window could have begun on January 1, 2013. See § 203(a)(3).
table litigation accompanying them) will turn, in part, on the outcome of the work made for hire analysis and what it means to be an employee versus an independent contractor, because termination of transfer rights do not exist for works made for hire.\textsuperscript{322} Instead, an employer whose employee created a copyrightable work will be the author and can enjoy the copyright for the duration without the fear of losing the copyright thirty-five to forty years later. To be sure, terminations of transfers under section 203 apply to all types of copyrightable works. But the looming litigation involving terminations of transfers of sound recordings is a good case study for analyzing how the \textit{Reid} factors should be analyzed and what the likely outcome is in light of the factors’ relative importance.

To appreciate the work made for hire analysis of sound recordings, it is necessary to understand what sound recordings are and the potential authors laying claim to those copyrights. Sound recordings are “works that result from the fixation of a series of musical, spoken, or other sounds . . . regardless of the nature of the material objects . . . in which they are embodied.”\textsuperscript{323} Sound recordings are to be distinguished from musical works, which are the underlying composition and lyrics.\textsuperscript{324} An artist who sings and records a song written by someone else has created a copyrightable sound recording but has no copyright interest in the underlying composition.\textsuperscript{325}

There are three major players in the sound recording industry who are likely to make a claim as being an author of a particular sound recording. First are the recording artists themselves. The artists are the individuals in the recording studio and creating the sounds by playing the instruments and singing the lyrics. Artists could be solo artists such as Billy Joel and Madonna or groups like Aerosmith and Bon Jovi. The House Report accompanying the Sound Recording Act of 1971\textsuperscript{326} suggests that recording artists can be authors of sound recordings.\textsuperscript{327} The Copyright Office does the same.\textsuperscript{328}

\begin{itemize}
\item \textsuperscript{322} See 17 U.S.C. § 203(a) (“In the case of any work other than a work made for hire, [a transfer of rights] is subject to termination . . . .”) (emphasis added)).
\item \textsuperscript{323} \textit{Id.} § 101 (defining “sound recordings”).
\item \textsuperscript{324} See Jessica L. Bagdanov, Comment, \textit{Internet Radio Disparity: The Need for Greater Equity in the Copyright Royalty Payment Structure}, 14 CHAP. L. REV. 135, 139 (2010).
\item \textsuperscript{325} See Brian Day, Note, \textit{The Super Brawl: The History and Future of the Sound Recording Performance Right}, 16 MICH. TELECOMM. & TECH. L. REV. 179, 182-83 (2009).
\item \textsuperscript{326} The legislation in the Sound Recording Act of 1971 gave federal copyright protection to sound recordings. Before then, the only protection afforded was state common law copyright protection. Daniel Gould, \textit{Time’s Up: Copyright Termination, Work-For-Hire and the Recording Industry}, 31 COLUM. J.L. & ARTS 91, 97-98 (2007).
\end{itemize}
Second are the producers. Producers come in all forms, but many of the most successful are those who are the driving force behind the recording; they are the ones with a vision of the recording and orchestrate everything from the lead vocals to instrumental solos to background vocals. Producers bring their experience with songwriting and arranging, musical performance, and recording to the table along with their musical philosophy, knowledge of the music business, and rapport with artists to create the sound eventually heard by the public. Just as with recording artists, the House Report and Copyright Office suggest that producers may have a claim to authorship of sound recordings.

Third are the record companies. The role of the record company has changed over time. Initially, the record companies exercised a great deal of control over the creative process. But beginning in the 1970s, record companies narrowed their focus to manufacturing and promoting the sale of records. Nonetheless, record companies have not really abdicated all control over the creation of sound recordings and have a strong interest in claiming these recordings as works made for hire. In fact, nearly all recording contracts between record companies, artists, and producers contain clauses stating that the sound recordings are works made for hire and are owned by the record companies. Unfortunately for the record companies, merely

---

329. Richard James Burgess, The Art of Music Production 60-61 (Oxford University Press 4th ed. 2013) (describing how producers “can influence the intuitiveness of the arrangement so it flows naturally, is appealing on a first listen, creates a desire to be heard again, and holds the listener’s interest through repeated plays”); M. William Krasilovsky & Sidney Shemel, This Business of Music 37 (Billboard Books 8th ed. 2000).


332. See Menell & Nimmer, supra note 330, at 1 (noting that before the 1960s, the A&R departments at the record companies “nurtured the talent by locating, writing, and arranging music, recruiting studio musicians, and arranging recording sessions, often in label-owned or affiliated studios”).


334. See infra notes 401-14 and accompanying text.

335. Henslee & Henslee, supra note 321, at 697 (“Record companies prefer for sound recordings to be considered under the ‘work-made-for-hire’ doctrine because it prevents this right of termination and recapture.”).

declaring a work to be a work made for hire does not necessarily make it so.\textsuperscript{337} This is because such statements are only relevant for specially commissioned works made for hire, not those falling under the employee within the scope of employment provision.\textsuperscript{338}

As a result, an analysis of the \textit{Reid} factors is necessary to determine whether artists and producers are employees of the record companies or independent contractors and hence whether the sound recordings are works made for hire.\textsuperscript{339} The remainder of this part analyzes the \textit{Reid} factors in light of common practices in the music business. Then, using the continuum of importance, this part determines whether artists and producers are likely to be deemed employees and provides an in-depth roadmap for litigants and judges to analyze this issue in the upcoming cases. That being said, a word of caution is in order. Like with most general rules, there are exceptions. Contracts and practices in the record industry are no different. Every recording contract is not the same and what may be a general practice might not apply in a particular case.

(a) \textit{Group #1}

Starting with the most important group of factors, the first factor to consider is the tax treatment. This factor likely weighs in favor of
artists and producers, because record companies “rarely withhold income taxes or contribute to social security . . .”340

The second factor is whether employee benefits are provided. This factor is a tossup with respect to artists, but it definitely weighs in favor of producers. As a general rule, record companies do not provide health insurance, dental insurance, or retirement funds.341 Nonetheless, the two major artist unions—AFTRA and AFM—provide health and retirement funds for their members.342 Although not directly providing these traditional benefits, record companies are required to contribute to these unions’ health and retirement funds as part of their agreements with the unions.343 As a result, although not directly paying for employee benefits, the record companies are indirectly providing them to those who qualify. Similar agreements for the funding and provision of benefits do not exist for producers.344 In sum, record companies have a colorable argument that they provide benefits to artists, but not for producers. This factor could go either way with respect to artists, but it certainly weighs in favor of producers.

The last factor in this group is the method of payment. Although record companies have a little room to argue this factor weighs in their favor, it overwhelmingly favors artists and producers being treated as independent contractors. Artists and producers are compensated identically in two ways—advances and royalties. Artists are also compensated in a way producers are not—union scale.345 Artists and producers receive advances as part of a recording fund or are given separate cash advances that are not tied to the recording fund.346 Under a recording fund arrangement, the artist or producer

340. KRASILOVSKY & SHEMEL, supra note 329, at 203.
342. KRASILOVSKY & SHEMEL, supra note 329, at 64 (“The AFTRA Health and Retirement Funds provide medical coverage and retirement benefits for eligible AFTRA members.”); MCPHERSON, supra note 341; see BEVERLY HILLS BAR ASS’N, supra note 336, at 229; Menell & Nimmer, supra note 330, at 5.
343. KRASILOVSKY & SHEMEL, supra note 329, at 60, 64.
344. E-mail from Jeff Slattery, Assistant Professor of Law and Arts & Entertainment Law Project Director at Thomas Jefferson School of Law, to author (Jan. 20, 2014, 10:25 PM EST) (on file with author) (confirming that no producers’ union exists that provides similar benefits to authors and that agreements between producers and record companies do not provide for such employee benefits). Professor Slattery also notes that many artists do not receive health and retirement benefits under their agreements with labels, union membership, or otherwise and thus are similarly situated as producers. E-mail from Jeff Slattery, Assistant Professor of Law and Arts & Entertainment Law Project Director at Thomas Jefferson School of Law, to author (Jan. 22, 2014, 7:24 PM EST) (on file with author).
345. Menell & Nimmer, supra note 330, at 5 (“Throughout the post 1978 period, artists have traditionally been paid through advances against future royalties, although studio recording time may be credited at union scale.”).
346. BURGESS, supra 329, at 171; KRASILOVSKY & SHEMEL, supra note 329, at 203; MCPHERSON, supra note 341, at 61.
pays for the recording costs and if any money remains at the close of production, the artist or producer keeps the rest.\(^{347}\) Any advances or recording funds are paid in installments with a percentage paid before commencement of the recording, and the remainder of the payments are paid at various benchmarks along the way and upon delivery and acceptance of the masters.\(^{348}\) Compensation via an advance suggests that artist and producers are independent contractors, as payment by the job, instead of by the hour, is indicative of an independent contractor relationship.\(^{349}\)

Artists and producers are also paid a royalty based on sales of the album.\(^{350}\) These royalties are not paid to the artist and producer until the record company has recouped its recording costs, including any advances paid to the artist and producer.\(^{351}\) Once recouped, record companies generally pay artists’ and producers’ royalties twice a year, although some do so quarterly and others annually.\(^{352}\) Like with advances, the payment of royalties is not like an hourly wage or fixed salary paid to employees at frequent intervals. The payments cover a several-month period and the amount varies depending on the success of the album. As a result, this method of payment also favors artists and producers being classified as independent contractors.

Finally, artists may be compensated at union scale.\(^{353}\) For singers, as opposed to musicians, AFTRA provides minimum rates that artists must be compensated at, even if the artists are entitled to royalties.\(^{354}\) Union scale under AFTRA is calculated as the greater of a per-hour amount or per-side (per song) amount.\(^{355}\) However, it should be noted that these payments are capped at three times the minimum scale per side.\(^{356}\) With respect to payment of artists at union scale as calculated by the hour, this method of payment looks more akin to an hourly-wage arrangement for an employee.\(^{357}\) Nonetheless, the fact that record companies pay this amount at all has more to do with their agreement with the union rather than the relationship to the

\(^{347}\) McPherson, supra note 341, at 61; Schulenberg, supra note 336, at 108.
\(^{348}\) See Burgess, supra 329, at 171; Schulenberg, supra note 336, at 211-12.
\(^{349}\) Restatement (Second) of Agency § 220 cmt. j (1958).
\(^{350}\) Krasilovsky & Shemel, supra note 329, at 39-40 (producers); id. at 203 (“[R]ecording artists] rely on royalty compensation, which is contingent on success.”).
\(^{351}\) Id. at 23 (“Artists generally do not receive any royalties until the record company has recovered all of its recording costs incurred for the artist’s records.”).
\(^{352}\) Schulenberg, supra note 336, at 121.
\(^{353}\) Menell & Nimmer, supra note 330, at 5.
\(^{354}\) Beverly Hills Bar Ass’n, supra note 336, at 230.
\(^{355}\) Id.; Krasilovsky & Shemel, supra note 329, at 17 (“A side customarily consists of a single composition with a minimum duration of 2 ½ minutes.”).
\(^{356}\) Beverly Hills Bar Ass’n, supra note 336, at 230.
\(^{357}\) See Restatement (Second) of Agency § 220 cmt. j (1958).
artist, which undercuts the force of this argument.\textsuperscript{358} But if payments are calculated by the per-side method, they appear to be payments by the job and artists are likely to be deemed independent contractors.\textsuperscript{359}

In sum, the method of payment factor overwhelmingly favors artists and producers being treated as independent contractors.

\textit{(b) Group #2}

Moving on to the second most important group of factors, the first factor to consider is whether the hiring party has the right to assign additional projects to the hired party. This factor probably favors artists being treated as employees, but producers as independent contractors. Most recording contracts are structured so the artist is required to create one album, but the record company is also given a series of options to extend the agreement for several more albums.\textsuperscript{360} The number of options the record company has depends on the bargaining power of the artist but five or six options are common.\textsuperscript{361} Unlike artists, agreements between producers and record companies do not give the record company options to require the producer to produce another album.\textsuperscript{362}

The second factor in this group is the skill required. This factor likely weighs in favor of artists and producers being independent contractors. The record companies expend a tremendous amount of energy trying to find which artists to sign and which to pass on.\textsuperscript{363} Not every artist can fill the void in the market the record company is looking to capture and performing that sound can be quite a unique skill.\textsuperscript{364} The record companies seem to acknowledge this unique skill when they include provisions in the recording contracts that the artists are of a special and unique character that gives them peculiar

\textsuperscript{358} This same argument could undercut record companies’ positions with respect to artists’ benefits, which is already an uncertain proposition. \textit{See supra} notes 341-44 and accompanying text.

\textsuperscript{359} \textit{See Restatement (Second) of Agency § 220 cmt. j (1958)}.

\textsuperscript{360} \textit{Krasilovsky & Shemel, supra} note 329, at 203 (“Many recording contracts call for additional recordings at the option of the record company.”).

\textsuperscript{361} \textit{McPherson, supra} note 341, at 59.

\textsuperscript{362} \textit{See Burgess, supra} note 329, at 168 (“Producer contracts rarely extend beyond one album . . .”); \textit{Schulenburg, supra} note 336, at 210-11 (describing how producers may be given a right to produce subsequent albums if a certain number of copies are sold).

\textsuperscript{363} \textit{Beverly Hills Bar Ass’n, supra} note 336, at 322-23; \textit{see also} Peter Muller, \textit{The Music Business – A Legal Perspective: Music and Live Performances} 79 (Quorum Books 1994) (describing what A&R departments do); \textit{Schulenburg, supra} note 336, at 21 (describing the major function of A&R departments at record companies as discovering new talent); Alan H. Siegel, \textit{Breakin’ In to the Music Business} 25, 91-92 (Cherry Lane Books 1986) (illustrating enormous amount of time and energy A&R departments spend trying to find good artists).

\textsuperscript{364} \textit{See Krasilovsky & Shemel, supra} note 329, at 203.
value.\(^{365}\) Although drafted to allow record companies to obtain injunctive relief if artists try to record for other record companies during their contracts, these provisions may likely be used to hoist the record companies by their own petards. For producers, creating a sound recording is a highly-skilled job in that the producer not only needs to have a grand vision for what the album or record will sound like but also frequently makes musical suggestions to achieve that sound.\(^{366}\) As with artists, the contracts between record companies and producers contain clauses providing that the producers and their skills are special and unique.\(^{367}\)

The final factor in this group looks at the source of the instrumentalities and tools. This factor likely favors artists and producers or is neutral. Although record companies sometimes own their own recording studios, it is rare that artists actually record there.\(^{368}\) If the recording is done at an independent studio, this factor would be neutral. However, sometimes the producer or even the artist owns the recording studio and equipment.\(^{369}\) In such a case, this factor could weigh in favor of the artists or producer being an independent contractor. As far as providing the instruments used during the recording sessions, these are normally owned by the artists and are sometimes purchased or paid off using the advance received from the record company.\(^{370}\)

\((c)\) **Group #3**

The third group of factors along the continuum of importance also has three factors. The first factor in this group is the extent of the hired party’s discretion over when and how long to work. Who this factor benefits may very well depend on the stature of the artist or producer. Brand new artists and producers are given little control over

---

365. **Beverly Hills Bar Ass’n, supra** note 336, at 358.

366. See **Burgess, supra** note 329, at 27, 182 (“Like all artistic endeavors, the successful pursuit of music production requires the development of practical skills.” Creating hit singles and albums “requires skill and experience.”); **Garofalo, supra** note 333, at 223 (“Overdubbing, layering, mixing, and the addition of special effects such as reverb, equalization, and compression removed increasingly important aspects of the creative process from the studio performance and located them in the control room, elevating the producer to an artistic status equal to that of the musicians he produced.”).

367. **Beverly Hills Bar Ass’n, supra** note 336, at 397-98.

368. **Krasilovsky & Sheemel, supra** note 329, at 203.

369. Nimmer & Menell, **supra** note 339, at 399 (referring to testimony by Sheryl Crow that she owns her own recording equipment).

issues such as the time for recording.\textsuperscript{371} But over time, artists and producers can secure control over this aspect of their recordings.\textsuperscript{372}

The second factor in this group is whether the work is part of the regular business of the hiring party. This factor certainly favors the record companies.\textsuperscript{373} Record companies are, and always have been, in the business of creating or acquiring rights to sound recordings and having those recordings distributed.\textsuperscript{374} Recording the albums that will eventually be distributed falls perfectly in line with these business practices.

The final factor in this group is the duration of the relationship between the parties. Artists and producers will probably be treated differently with respect to this factor. As mentioned earlier, recording contracts with artists are typically for a certain number of albums, and the record company has several options to extend the relationship.\textsuperscript{375} As a result, the exact duration of the relationship can be difficult to determine,\textsuperscript{376} but typically lasts for a period of several years,\textsuperscript{377} although usually not more than five to seven years.\textsuperscript{378} These long durations weigh in favor of artists being deemed employees.

\textsuperscript{371} See SIEGEL, supra note 363, at 111 (“In the basic beginner situation, the record company quite understandably seeks to maintain as much control as possible. Their experience has shown them that too often when they have relinquished control, disaster results); BEVERLY HILLS BAR ASS’N, supra note 336, at 331 (sample recording contract stating “The masters recorded hereunder by Artist shall be recorded in a recording studio selected or approved by Company at such times as Company may designate or approve.”); id. at 340 (another sample recording contract providing that the artist’s recording budget must include “the dates of recording and mixing . . .” and that the artist may begin recording after the record company has given its written approval).

\textsuperscript{372} KRASILOVSKY & SHEMEL, supra note 329, at 203 (“A record company’s control is likely limited to choosing producers and material to record.”); SIEGEL, supra note 363, at 121-23.

\textsuperscript{373} KRASILOVSKY & SHEMEL, supra note 329, at 203.

\textsuperscript{374} AL KOHN & BOB KOHN, KOHN ON MUSIC LICENSING 1369 (3d ed. Aspen Pub. 2002) (record companies typically own the rights to sound recordings); NEW YORK STATE BAR ASS’N, ENTERTAINMENT LAW 29 (Howard Siegel ed., 3d ed. 2004) (“Typically, the record company acquires all right, title and interest in and to the masters . . . .”); SCHULENBERG, supra note 336, at 149 (“Without the rights [to the sound recordings], the record company has nothing . . . .”); see DONALD S. PASSMAN, ALL YOU NEED TO KNOW ABOUT THE MUSIC BUSINESS 65-66, 68 (8th ed. 2012) (describing distribution by major and independent record companies).

\textsuperscript{375} See KRASILOVSKY & SHEMEL, supra note 329, at 14 (“The standard contract runs for a term based on the delivery of a minimum number of recordings (usually enough to constitute one album) plus a period of time thereafter (usually 9 months). The record company may have an option to extend the agreement for one or more successive albums.”); supra notes 360-61 and accompanying text.

\textsuperscript{376} SIEGEL, supra note 363, at 105.

\textsuperscript{377} KRASILOVSKY & SHEMEL, supra note 329, at 203.

\textsuperscript{378} BEVERLY HILLS BAR ASS’N, supra note 336, at 223 (describing that the AFM by-laws do not allow musicians to enter into personal service contracts for more than five years without approval); SCHULENBERG, supra note 336, at 28 (“In general, the total con-
Producers, on the other hand, are not typically signed to agreements giving the record companies options to have producers work on subsequent albums. Instead, a producer’s working relationship with the record company is based upon the success of earlier recordings. Producers are almost always hired on a project basis rather than a number of years. In fact, sometimes producers are hired to work on a single song rather than an entire album. As a result, the duration factor weighs in favor of producers being treated as independent contractors.

(d) Group #4

Moving on to the less important end of the continuum, there are two factors to consider. The first factor is the location of the work. This factor likely weighs in favor of artists and producers as independent contractors. Although record companies may have their own in-house studios and prefer that they be used, it is rare for artists to record at those studios. Instead, the record companies try to include a provision in their contracts giving the record company the right of final approval of the recording studio. In practice, the record companies will approve any legitimate studio. Given that very few recordings occur at the record companies’ studios, that the choice of the recording studio is made by artists and producers, and that the record companies rarely exercise their ability to veto a location, this factor probably favors artists and producers as independent contractors.

The second factor in this group is the hired party’s role in hiring and paying assistants. This factor probably weighs in favor of artists and, to a lesser extent, producers being independent contractors, but may depend on the use of a recording fund and recoupment. In terms of selecting those who contribute to the recording, the artists and producers typically choose which engineers, non-featured musicians,
and non-featured vocalists to hire. The selection of the producer is a bit more involved. To the record company, selection of the producer is one of the more important provisions of the recording contract. These provisions usually state that the producer will be mutually selected by the artist and record company. In practice, however, the artist selects the producer, the record company consults and generally defers to the artist unless the record company has had a bad experience with the artist’s choice. Thus, it appears that artists and producers play a large role in hiring assistants.

Paying for those assistants is a bit more complicated. Today, artists pay for third parties out of a recording fund. Under this approach, the recording costs and artist’s advance are combined into a single fund, and the artist is responsible for paying all recording costs out of this fund. Any money leftover is the artist’s to keep as an advance. As a result of this payment scheme, recording contracts include clauses specifying that the artist is solely responsible for paying all third party charges incurred in the production of the record. Because the artist’s advance and recording costs are combined into one fund that is delivered to the artist to administer, it appears that this factor weighs in favor of the artist being an independent contractor.

Although the recording fund is more common today, it was common for the artist (especially a new artist) to be paid a separate advance and for the recording company to pay the recording costs. Although this arrangement suggests that the record companies paid for assistants, it is important to remember that all of the recording

---

388. Id.; Nimmer & Menell, supra note 339, at 399 (describing Sheryl Crow’s testimony that she decides what musicians she wants to perform on each song, what engineers she wants to hire to implement her “sonic vision”).

389. BEVERLY HILLS BAR ASS’N, supra note 336, at 332.

390. Id.

391. Id.

392. KRASILOVSKY & SHEMEL, supra note 329, at 23-24; SIEGEL, supra note 363, at 112.

393. BEVERLY HILLS BAR ASS’N, supra note 336, at 338.

394. Id. ("[The recording fund] gives the artist an incentive to minimize the recording costs, since the unspent portion of the recording fund is paid to the artist as a cash advance."); JEFFREY BRABEC & TODD BRABEC, MUSIC, MONEY, AND SUCCESS 88 (1994) ("[I]f the artist’s expenses do not exceed the recording fund . . . the amount left over is kept by the artist and treated as an additional advance recoupable from the artist’s royalties.").

395. BEVERLY HILLS BAR ASS’N, supra note 336, at 341.

396. The recording fund is customarily distributed at three points: (1) upon commencement of recording, (2) upon completion of the basic tracks, and (3) upon delivery of the finished master to the record company. KRASILOVSKY & SHEMEL, supra note 329, at 23-24.

397. BEVERLY HILLS BAR ASS’N, supra note 336, at 338; BRABEC & BRABEC, supra note 394, at 88.
costs are treated as advances to the artist. As a result, these costs are recoupable from the artist’s royalties. Because of this, if the artist earns a sufficient amount of royalties, it is the artist, not the record company, who pays for the recording costs. Of course, if the artist is never successful, then the record company will not recoup these costs and will end up paying for them. Who ultimately pays for the assistants under the older method of payment depends on whether the artist’s album is a commercial success.

In sum, this factor probably weighs in favor of artists and, to a lesser extent, producers being independent contractors. Regardless of how these assistants are paid, the artist and producer choose who works on the recording. When a recording fund is used, the case for independent contractor is strengthened. However, if the record company did not use a recording fund and the costs are not recouped, then the record company will have a stronger argument that this factor weighs in favor of the artist and producer being employees.

(e) Group #5

Concluding with the least important group of factors, the first factor to consider is the right to control the manner and means by which the product is accomplished. Whether this factor weighs in favor of artists and producers being independent contractors or employees is difficult to determine. Prior to the 1970s, record companies exercised a large amount of creative control over the recording process, but since then, artists and producers have moved away from an in-house creative process and are exercising more creative control in recording. Today, for newer artists and producers, the record companies may have language in the recording contracts requiring the artist and producer to submit written recording budgets specifying who the producer is and the financial agreement between the artist and producer, the songs to be recorded, the accompaniment and arrangement of the recordings, any recording fees that will exceed union scale, the dates and locations of the recordings and mixing, and the estimated costs of these sessions. Despite this contract language, in practice,
written recording budgets are not always submitted or followed. More experienced artists and producers may, however, secure more control over the creative process, including the selection of music, recording location and time, final sound mix, and recording budgets. As a general matter, the record companies may have a better chance at artists and producers being employees under the right to control factor when the artists and producers are inexperienced. But the amount of control over the creative process wanes as artists and producers get more experience.

Specific components of recording that are important in the creative process are controlling music selection, monitoring what takes place during recording sessions, and selecting where the recording occurs. For music selection, the standard recording contract provides that the record company will select the music to be recorded. But in practice, the record company gives the artist a say in the decision. It is common today for the record company to only reserve a right of approval of the music selection. This is especially true for more seasoned artists. That said, there are actually very few disputes about song choice.

With respect to where the recording occurs, as discussed in connection with the work location, record companies try to include clauses giving the record company approval rights of the recording studio. In practice, the producer will normally choose the studio to work in and the record companies will approve any legitimate studio. In addition to where the recording occurs, recording contracts typically contain a provision that the record company has the right to have a representative present to supervise the recording session.

403. Id.
404. See SIEGEL, supra note 363, at 122-23.
407. BEVERLY HILLS BAR ASS‘N, supra note 336, at 332.
408. KRASILOVSKY & SHEMEL, supra note 329, at 19.
409. BEVERLY HILLS BAR ASS‘N, supra note 336, at 332.
410. Id. at 331 (“The masters recorded hereunder by Artist shall be recorded in a recording studio selected or approved by Company at such times as Company may designate or approve.”); MULLER, supra note 363, at 82.
411. WEISSMAN, supra note 370, at 48.
412. BEVERLY HILLS BAR ASS‘N, supra note 336, at 332 (“Generally, the record company will approve any recording studio the artist or the artist’s producer wishes to use, so long as it has satisfactory recording equipment and the studio’s rates fit within the recording budget.”); McPherson, supra note 341, at 83.
413. SIEGEL, supra note 363, at 112-13.
Nonetheless, it is very rare for record companies to actually send a representative to the studio to supervise.\textsuperscript{414}

Although it appears that record companies actually exercise very little control over the creative process, they do reserve the right to do so in the recording contracts they enter into with artists and producers. As this factor really focuses on the right to control instead of actual control, which was rejected by the Supreme Court in \textit{Reid}, it is likely that this factor will weigh in favor of artists and producers being considered employees.

The second factor in this group is the label the hired and hiring party use to describe the hired party. This factor certainly weighs in favor of artists and producers being deemed employees. Nearly every recording contract written since 1978 states that the works created by artists and producers are works made for hire.\textsuperscript{415} Although many of the terms of a recording contract are negotiable, this one is not.\textsuperscript{416}

The final factor in the least important group of factors is whether the hiring party is in business. Record companies are obviously in business.\textsuperscript{417} This factor clearly weighs in favor of the artists and producers being employees.

\textit{(f) Conclusions for the Music Industry}

Table 6, below, summarizes the analysis of the \textit{Reid} factors as applied to sound recordings in the music industry. This table sorts the factors by order of importance. A checkmark indicates this factor likely weighs in favor of that party’s favored position (i.e. independent contractor for artist and producers, employees for record companies). A question mark indicates that it is unclear which way that factor applies.

\textsuperscript{414. \textit{Id.}}

\textsuperscript{415. Nimmer & Nimmer, supra note 75, at § 5.03[B][2][a][iii][II] ("[V]irtually all contracts that artists signed with record companies from 1972 onwards have contained acknowledgements of their contributions constituting works for hire."); Nimmer & Menell, supra note 339, at 396; see also McPherson, supra note 341, at 90; Schulenberg, supra note 336, at 148, 150.}

\textsuperscript{416. See Beverly Hills Bar Ass’n, supra note 336, at 390.}

\textsuperscript{417. Krasilovsky & Shemel, supra note 329, at 203.}
As illustrated in Table 6, it is nearly certain that producers will be treated as independent contractors. Five, and possibly six, of the most important factors weigh in favor of producers as independent contractors. In fact, the only factors suggesting producers would be employees are the part of the regular business of the hiring party and the three least important factors on the continuum.

The case for artists is more difficult. Nonetheless, it is more likely that artists will be deemed independent contractors. Of the three most important factors, two clearly weigh in favor of artists as independent contractors. The other factor (employee benefits) is unclear. As described earlier, when a party has two of these three factors in its favor, the courts find in favor of that party eighty-seven percent of the time.418 Expanding this to the next group of factors, one favors the record companies, one favors the artists, and the remaining factor either favors the artists or is neutral. If a court were to rule that either the employee benefits or source of the instrumentalities factor favored the artists, then a majority of the top six factors would favor the artists. As described earlier, when a party has a majority of these six factors in its favor, the courts find in favor of that party ninety-one percent of the time.419

---

418. See supra Figure 2.
419. See supra Figure 2.
companies when the artist is new and is split two to one when the artist is established. This group is helpful to the record companies, but they face an uphill battle given that the more important factors favor artists. The fourth group of factors favors artists, which tips the scales slightly more towards artists being considered independent contractors. However, the least important factors all support the record companies. Of course, these are the least important factors, so their utility to the record company is of little value. Given the results of the most and second-most important groups of factors, this will likely carry the day and artists will be deemed independent contractors. That said, the record companies have a colorable argument that artists should be treated as employees, and this argument should not be considered futile.420 Hopefully, the results of this study and the analysis above provide a useful roadmap for litigants and judges navigating the upcoming termination of transfer cases.

3. Business Planning

In addition to assisting litigators to evaluate and present their cases, the results of this study will aid business planners in structuring relationships between hiring and hired parties. For example, if a hiring party seeks initial ownership of the copyright, it will be best advised to withhold income taxes and issue a W2 rather than a 1099, provide employee benefits such as life and health insurance, and pay the hired party at regularly intervals. To further ensure a work made for hire result, the hiring party could also include a provision in the agreement that reserves the right for the hiring party to assign additional projects to the hired party. And although difficult to control in some circumstances, the hiring party should provide as many of the tools as possible. Structuring the relationship this way forces the three most important factors and one or two of the factors in the second group to weigh in favor of employee status. This should all but assure a conclusion that the hired party will be an employee.

Likewise, for attorneys representing hired parties who would like to retain initial ownership—to take advantage of the termination of transfer provisions or to further exploit the copyright—they should insist on the hiring party not withholding taxes and issuing a 1099, refuse insurance coverage, and demand payment upon completion of projects or portions of projects. Submitting invoices to the hiring party upon completion would be a wise practice to adopt. Moreover, resisting a provision to accept additional projects and actually refusing additional projects until a new agreement is established for a new

420. See Mark H. Jaffe, supra note 337, at 165 ("[I]t would be futile for a record company to press this argument in litigation.").
project will place the hired party in a strong position to argue that she is an independent contractor.

Of course, negotiating these terms may heavily depend on bargaining power. A weak hired party may not be able to demand that no additional projects be assigned. Likewise, a weak hiring party may not be able to insist that the hired party work on the hiring party’s premises. But knowing where the various factors lie on the continuum may help the attorney for the weaker party focus on specific factors that the stronger party may be willing to budge on. If choosing between asking the stronger party to capitulate with respect to a provision about where the hired party will physically work and a provision concerning the hours of the day the hired party must work, the attorney for the weaker party should focus negotiation efforts on the working hours provision, because it is more important than the work location factor.

V. CONCLUSION

Drawing a distinction between employees and independent contractors in the work made for hire doctrine is a challenging endeavor, but one that strikes at the heart of many copyright disputes. After a quarter century of cases applying and interpreting the Supreme Court’s multifactor test from CCNV v. Reid, this Article gives the first comprehensive study of this multifactor test and answers the question of which factors are the most and least important in these analyses. Some of the results are surprising while other results are expected. But regardless of how these results conform to our expectations, they will prove useful to the bench and bar involved in copyright litigation and for business planning attorneys advising their clients how to accomplish their copyright ownership goals.
I. INTRODUCTION

A primary focus of American grand strategy in foreign affairs is to promote democratic, prosperous societies that are inhospitable to political extremism.\(^1\) At the same time, the United States regularly sends unmanned aerial vehicles (UAVs, or “drones”) to kill suspected terrorists in multiple countries\(^2\) and strictly enforces a statutory prohibition against providing terrorists with “material support.”\(^3\) In theory, these two approaches—one preventative and one reactive—should complement each other in the service of widely shared global interests. In practice, however, the increasingly dominant role of reactive, zero-tolerance antiterrorism policies is both crowding out preventative American public diplomacy and expanding the scope of the executive branch’s national security authority.

President Barack Obama kept his campaign promises to refocus American military power on al-Qaeda after the wars in Afghanistan


and Iraq, but he has not rejected the “unchecked presidential power” for which he criticized his predecessor. Despite the ambiguous legal basis of targeted killing and its inevitable collateral damage, the Obama Administration has standardized the use of aerial drone strikes and dramatically increased the number of countries in which they occur. Meanwhile, executive agencies have interpreted the material support statutes so broadly that ordinary civilians and established charities are easily tainted by incidental interactions with such groups. Both drone strikes and the material support prohibitions share a failure to adequately distinguish terrorists from the communities in which they live and operate.

This Comment advances the premise that, when paired with the President’s political incentives, the executive’s constitutional duty to defend the United States from attack inevitably leads to an overbroad conception of self-defense and a maximalist approach to executive power. In a government of checks and balances, the corresponding duty of the other two branches is to prevent that result. But although Congress has explicit constitutional war powers and has expressed some intent to exercise them, the special threat of terrorism has inspired extreme legislative deference to the Commander-in-Chief. The judiciary, for its own part, has consistently denied its authority to properly balance national security interests against individual civil liberties. Evidence is mounting that this double deference is undermining the rule of law and harming long-term foreign policy objectives.

Instead of relying on executive restraint or stricter judicial review in this area, I suggest that those of us who are troubled by these trends must look primarily to Congress to reverse them. The legislature has clear constitutional authority to regulate the use of military force, to define the limits of military counterterrorism operations and set other foreign policy objectives, and to give courts better statutory parameters to increase their respective oversight of the executive in its pursuit of national security. Part II lays out the basic legal framework for initiating U.S. military action, tracing the roots of deference to the executive and describing current antiterrorism policies. Part III surveys some key consequences of antiterrorism absolutism and unchecked executive power for both domestic and foreign affairs. Finally, Part IV proposes that Congress should begin to assert its


5. See infra Part II.B.2.


7. See infra Part II.A.2.
check on national security policy by revisiting three statutes: the outdated Authorization for Use of Military Force,8 the prohibition on providing material support to terrorist groups,9 and the Foreign Assistance Act.10 Amending these laws would give the legislature more opportunities to balance executive power and develop better and more accountable policies.

II. The Balance of National Security Authority

Constitutional provisions relating to the use of military force offer minimal guidance as to the appropriate interbranch balance of power. This section contrasts the vague constitutional promise of a congressional check with the reality of its inadequacy, particularly for executive action where the threat of terrorism is apparent. It concludes that U.S. Presidents’ expansive interpretations of executive war powers—including those of President Obama—should not be at all surprising, given the executive’s vague legal parameters, political incentives, and lack of effective oversight.

A. Constitutional Powers

1. Text and Early Debates

When the delegates to the 1787 Constitutional Convention gathered in Philadelphia, the vesting of war-making authority outside of a sovereign executive was unprecedented; yet the drafters struggled to limit the American executive’s ability to deploy military force.11 The only constitutional provision explicitly granting such authority is Article II, section 2, which makes the President “the Commander-in-Chief of the Army and Navy of the United States . . . .”12 Article II also gives the President influence over other foreign policy matters, such as the power to “receive [foreign] Ambassadors”13 and to make treaties and appoint American ambassadors, with the advice and consent of the Senate.14 Finally, the executive must “take Care that the Laws be faithfully executed . . . .”15

11. Many were fearful of an executive with control of a standing army. See, e.g., THE FEDERALIST NO. 26 (Alexander Hamilton).
13. Id. § 3.
14. Id. § 2, cl. 2.
15. Id. § 3.
In contrast to Article II’s unconditional grant of executive power, Article I limits congressional authority to the “legislative Powers herein granted.” 16 However, the powers enumerated in section 8 establish broad legislative control over military and foreign policy. Congress has the authority to “declare War, grant Letters of Marque and Reprisal, and make Rules concerning Captures on Land and Water.” 17 It can also “raise and support Armies,” provided that “no Appropriation of Money to that Use shall be for a longer Term than two Years;” similarly, it can “provide and maintain a Navy.” 18 Related powers include: “To make Rules for the Government and Regulation of the land and naval Forces; To provide for calling forth the Militia to execute the Laws of the Union, suppress Insurrections and repel Invasions; [and] To provide for organizing, arming, and disciplining, the Militia.” 19 Section 8 also establishes significant legislative control over U.S. monetary policy and gives Congress the power “[t]o define and punish Piracies and Felonies committed on the high Seas, and Offences against the Law of Nations.” 20 Finally, Congress can “make all Laws which shall be necessary and proper for carrying into Execution” its other powers. 21

The inevitable tension between the legislative power to “declare war” and the executive’s commander-in-chief power emerged early, when President Washington unilaterally declared American neutrality in the French Revolutionary Wars. 22 James Madison protested that inherent in Congress’s authority to declare war was the inverse power to declare that the United States was not at war, that is, that the United States was neutral. 23 But Alexander Hamilton defended Washington’s declaration, invoking the unconditional and broadly defined vesting language of Article II. 24 Congress quickly mooted the

16. Id. § 1.
17. Id. § 8, cl. 11. The drafters at the Constitutional Convention argued as to whether Congress should have the power to “make” war or merely “declare” it, settling on the latter to avoid infringing on the Commander-in-Chief’s authority. 3 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, at 318-19 (M. Farrand ed., 1911).
19. Id. cl. 14-16.
20. Id. cl. 10.
21. Id. cl. 18.
22. See 32 THE WRITINGS OF GEORGE WASHINGTON 430-31 (John C. Fitzpatrick ed., 1939). Then Chief Justice John Jay refused to advise the President on the legality of his action. Although many praise this decision as a key foundation of an independent judiciary, the incident also marks the beginning of the courts’ long tradition of avoiding questions related to foreign affairs and military policy in particular. Id.
controversy by passing its own declaration of neutrality, but the episode revealed a deep and persistent uncertainty about the scope of executive power under Article II.

2. Judicial Deference to the Executive

Over the past century, the Supreme Court has effectively validated the Hamiltonian view of executive power. In its precedential analysis in United States v. Curtiss-Wright Export Corp., the Court considered whether a Joint Resolution of Congress was sufficient to prohibit weapons sales to certain foreign countries if the President had withdrawn a proclamation to the same effect. The Court first distinguished the federal government’s “external” powers from its domestic powers: while federal authority over American citizens had been delegated by the states, its foreign relations power had been inherited from the previous sovereign, Great Britain. Bracketing foreign relations authority in this way, the Court reasoned that such inherited authority was largely vested in the President:

[P]articipation in the exercise of the power is significantly limited. In this vast external realm, . . . the President alone has the power to speak or listen as a representative of the nation. . . . [H]e alone negotiates. Into the field of negotiation the Senate cannot intrude; and Congress itself is powerless to invade it. As Marshall said in his great argument of March 7, 1800, in the House of Representatives, “The President is the sole organ of the nation in its external relations, and its sole representative with foreign nations.”

Citing a Senate Foreign Relations Committee report from 1816 that advocated presidential dominance in foreign relations, the Court further decided that the President’s power in this area “does not require as a basis for its exercise an act of Congress.” Justice Sutherland’s opinion in Curtiss-Wright emphasized the pragmatic need for unity of national representation in foreign relations and the President’s superior informational resources.

The Curtiss-Wright analysis has been subject to thorough evisceration in the academy, with respect to both its extra-constitutional theory of federal foreign relations power and its expansive conception

25. See Act of June 5, 1794, ch. 50, 1 Stat. 381.
27. Id. at 316.
28. Id. at 319.
29. Id. at 320.
30. Id.
of executive authority in that field.\textsuperscript{31} Nevertheless, courts still cite the decision whenever they defer to the executive in matters of foreign policy—which is to say, frequently.\textsuperscript{32} Even though Justice Sutherland’s theory of executive dominance was dicta, other cases quickly reinforced it. Five months after \textit{Curtiss-Wright}, Justice Sutherland expanded on the theory of executive dominance in \textit{United States v. Belmont}, in which the Court held that “executive agreements”—negotiated with foreign entities without legislative approval—carry the same legal weight as Senate-approved treaties under the law.\textsuperscript{33} Later, in \textit{United States v. Pink}, the Court held that the President’s constitutional authority to negotiate treaties, and the need for credibility in such interactions, justified the executive’s power to unilaterally make foreign policy.\textsuperscript{34} G. Edward White sees \textit{Curtiss-Wright’s} conclusions about executive power as a tipping point in what had been an “incremental” expansion of the President’s authority.\textsuperscript{35}

In 1952, the Court temporarily halted the legal expansion of executive power with its decision in \textit{Youngstown Sheet & Tube v. Sawyer}.\textsuperscript{36} The case arose from President Harry Truman’s Executive Order to seize steel mills, which were incapacitated by a labor strike, to maintain U.S. steel production.\textsuperscript{37} Premising its decision on the then-ongoing Korean War, the Truman Administration argued that the seizure served the interests of national security and was authorized by the executive’s commander-in-chief and implied emergency powers.\textsuperscript{38} The Court rejected the seizure’s constitutional legitimacy, expressing profound discomfort with the appropriation of private property in the name of national security.\textsuperscript{39} Writing for the majority, Justice Black summarily concluded that the seizure was a legislative action; and as such, it was unavailable to the executive, even in times

\begin{itemize}
\item \textsuperscript{31} See, e.g., David M. Levitan, \textit{The Foreign Relations Power: An Analysis of Mr. Justice Sutherland’s Theory}, 55 \textit{Yale L.J.} 467, 494 (1946) (arguing that \textit{Curtiss-Wright’s} historical argument for an extra-constitutional foreign relations power was unfounded); G. Edward White, \textit{The Transformation of the Constitutional Regime of Foreign Relations}, 85 \textit{Va. L. Rev.} 1, 98-110 (1999) (calling the \textit{Curtiss-Wright} decision revisionist and radical due to its abandonment of constitutional limits).
\item \textsuperscript{33} 301 U.S. 324, 331-32 (1937).
\item \textsuperscript{34} 315 U.S. 203, 228-30 (1942).
\item \textsuperscript{35} White, supra note 31, at 146.
\item \textsuperscript{36} 343 U.S. 579 (1952).
\item \textsuperscript{37} Id. at 582-83.
\item \textsuperscript{38} Id. at 583-84. President Truman immediately reported his action to Congress and sent another report twelve days later. After Congress did not respond, steel producers challenged the seizure in federal court. Id.
\item \textsuperscript{39} Id. at 587.
\end{itemize}
of war or emergency.\textsuperscript{40} However, the complexity of the question inspired five concurrences, including one by Justice Jackson that eventually became the controlling opinion of the case.\textsuperscript{41}

Justice Jackson thought that legitimate executive action could be conceptually organized around congressional approval, or lack thereof.\textsuperscript{42} When the President acts with the “express or implied authorization of Congress,” Justice Jackson wrote, “his authority is at its maximum.”\textsuperscript{43} When the President acts against the will of Congress, he can rely only on his independent constitutional powers.\textsuperscript{44} Between these two extremes,

there is a zone of twilight in which [the President] and Congress may have concurrent authority, or in which its distribution is uncertain. Therefore, congressional inertia, indifference or quiescence may sometimes, at least as a practical matter, enable, if not invite, measures on independent presidential responsibility. In this area, any actual test of power is likely to depend on the imperatives of events and contemporary imponderables rather than on abstract theories of law.\textsuperscript{45}

Justice Jackson concluded that even though Congress was silent after President Truman’s Order, its past history of strict regulation of private property seizures meant that the executive branch could only rely on its own constitutional authority for legitimacy in that case.\textsuperscript{46} Justice Jackson took pains to encourage a “practical” assessment of the scope of the commander-in-chief power but emphasized the need to limit it:

[J]ust what authority goes with the name has plagued Presidential advisers who would not waive or narrow it by nonassertion yet cannot say where it begins or ends. It undoubtedly puts the Nation’s armed forces under Presidential command. . . . But no doctrine that the Court could promulgate would seem to me more sinister and alarming than that a President whose conduct of foreign affairs is so largely uncontrolled, and often even is unknown, can vastly enlarge his mastery over the internal affairs of the country by his own commitment of the Nation’s armed forces to some for-

\textsuperscript{40} Id. at 588.
\textsuperscript{41} Id. at 634 (Jackson, J., concurring); see Michael J. Turner, Comment, \textit{Fade to Black: The Formalization of Jackson’s Youngstown Taxonomy} by Hamdan and Medellin, 58 Am. U. L. Rev. 665, 674 (2009).
\textsuperscript{42} Youngstown, 343 U.S. at 654.
\textsuperscript{43} Id. at 635.
\textsuperscript{44} Id. at 637.
\textsuperscript{45} Id.
\textsuperscript{46} Id. at 640.
eign venture. . . . He has no monopoly of “war powers,” whatever they are.47

Justice Jackson was even less amenable to the government’s implied emergency powers argument. The Framers, he wrote,

knew what emergencies were, knew the pressures they engender for authoritative action, knew, too, how they afford a ready pretext for usurpation. We may also suspect that they suspected that emergency powers would tend to kindle emergencies. Aside from suspension of the privilege of the writ of habeas corpus in time of rebellion or invasion, when the public safety may require it, they made no express provision for exercise of extraordinary authority because of a crisis. I do not think we rightfully may so amend their work.48

Thus, Youngstown limited the scope of executive power and articulated an enduring analytical framework that demanded consideration of legislative intent. However, its opinions repeatedly highlighted the domestic character of the steel mills and, therefore, ultimately provided little insight into the content and reach of executive power in true matters of foreign policy. Had the Court accepted the Truman Administration’s framing of the seizures as a national security necessity, Curtiss-Wright and its progeny indicate that the government’s arguments would have inspired much more deference.

Tracing several cases through the twentieth century, Anthony Simones writes that

[b]y the 1970s, the specific facts that gave rise to Curtiss-Wright faded from the memory of many judges who sought to use it as a precedent for presidential domination of national security affairs. Largely forgotten was Justice Robert Jackson’s reminder that Curtiss-Wright “involved, not the question of the President’s power to act without congressional authority, but the question of his right to act under and in accord with an Act of Congress.”49

Simones notes that no cases in this line seem to conceive of any limit on implied presidential powers in the realm of foreign relations,50 a trend which contrasts sharply with Justice Jackson’s Youngstown concurrence.51 Thus, at least with respect to foreign affairs, the judi-

47. Id. at 641-42, 644.
48. Id. at 650.
49. Simones, supra note 32, at 419.
50. Id.
ciary effectively resolved in Hamilton’s favor the early debate about the scope of executive war powers by the time global terrorism became a major security threat.

3. The War Powers Resolution and Congressional Acquiescence

With little help from the judiciary, Congress struggled for influence over American use of force as military operations became less conventional during the Cold War. In 1950, President Truman sent U.S. troops to the Korean peninsula, pursuant to a United Nations Security Council Resolution, without seeking congressional approval.\(^\text{52}\) He explained that this “police action” was an American obligation under the U.N. Charter.\(^\text{53}\) Many legislators disputed Truman’s authority to commit U.S. troops to the Korean conflict, and Congress never explicitly approved American participation.\(^\text{54}\)

The meaning of Congress’ constitutional power to “declare” war was uncertain in the age of nuclear weapons, when such declarations became nonsensical. Did “declare” encompass authorizations for war, and if so, to what extent?\(^\text{55}\) Protracted U.S. military operations in Vietnam raised the questions of whether implied authorization could be found in, for example, appropriation of funds for executive action.\(^\text{56}\)


56. See Orlando v. Laird, 443 F.2d 1039, 1043 (2d Cir. 1971), cert. denied, 404 U.S. 869 (1971) (holding that Congress’ appropriation of funds for the Vietnam Conflict constituted sufficient indication of its approval). In its power of the purse, Congress faces a political dilemma: it has constitutional authority to express disapproval by withholding appropriations for military operations, but this measure potentially turns deployed U.S. service
and under what circumstances the executive could proceed even in the case of unequivocal legislative disapproval.\textsuperscript{57}

The debacle of Vietnam led Congress to pass the War Powers Resolution of 1973 over presidential veto. The most assertive exercise of its constitutional war powers to date provides:

(a) Congressional declaration. It is the purpose of this chapter to fulfill the intent of the framers of the Constitution of the United States and insure that the collective judgment of both the Congress and the President will apply to the introduction of United States Armed Forces into hostilities, or into situations where imminent involvement in hostilities is clearly indicated by the circumstances, and to the continued use of such forces in hostilities or in such situations.

(c) Presidential executive power as Commander-in-Chief; limitation. The constitutional powers of the President as Commander-in-Chief to introduce United States Armed Forces into hostilities, or into situations where imminent involvement in hostilities is clearly indicated by the circumstances, are exercised only pursuant to (1) a declaration of war, (2) specific statutory authorization, or (3) a national emergency created by attack upon the United States, its territories or possessions, or its armed forces.\textsuperscript{58}

The Resolution also requires the President to report to Congress within forty-eight hours upon introducing military forces to hostilities without prior legislative approval.\textsuperscript{59} It also requires the President to remove those forces within sixty days, unless Congress specifically authorizes continued involvement.\textsuperscript{60} Finally, it explicitly rejects the use of American military force with no legal authorization other than an international treaty, even a treaty approved by the Senate.\textsuperscript{61}

Despite these directives, the four decades following the Resolution’s passage have proven the law to be toothless.\textsuperscript{62} No President has

\textsuperscript{57} Id. at 1040.

\textsuperscript{58} 50 U.S.C. § 1541 (2012).

\textsuperscript{59} Id. § 1543(a)(3).

\textsuperscript{60} Id. § 1544(b).

\textsuperscript{61} Id. § 1547(a).

\textsuperscript{62} However, there is no shortage of academic suggestions for improving the Resolution. See, e.g., Geoffrey Corn, Triggering Congressional War Powers Notification: A Proposal to Reconcile Constitutional Practice with Operational Reality, 14 LEWIS & CLARK L. REV. 687, 692 (2010); Jonathan T. Menitove, Note, Once More Unto the Breach: American War Power and a Second Legislative Attempt to Ensure Congressional Input, 43 U. MICH. J.L. REFORM 773, 791 (2010).
acknowledged the statute’s validity or any obligation to comply with it. Its language essentially allows a President to make war for two months without legislative approval—a generous timeframe for many potential military objectives, considering the power and precision of modern weaponry. In the sixteen major deployments of American military force since the Resolution became law, only five received explicit congressional approval.

The executive branch did not publicly interpret its constitutional war powers until 2011, when the U.S. Department of Justice released a memorandum in anticipation of the Obama Administration’s intervention in Libya. It asserted the President’s independent authority and obligation to determine and pursue American national security interests, derived from the executive’s commander-in-chief power and Supreme Court precedent since Curtiss-Wright. It also recognized a need for legislative authorization but only in the case of “prolonged and substantial military engagements, typically involving exposure of U.S. military personnel to significant risk over a substantial period.” The memo concluded that the Libyan operation did not rise to a level requiring approval from Congress. The United States’ intervention in Libya ultimately exceeded the War Powers Resolution’s sixty-day benchmark without a formal challenge.

In essence, then, the executive’s current position is that limits on its own military powers are to be primarily self-imposed. Congress has not effectively challenged that assertion, although it has the constitutional authority to do so.

63. The two conflicts in Iraq and the war in Afghanistan were preceded by explicit approval; whereas Lebanon and Somalia received approval after the fact. CURTIS A. BRADLEY & JACK L. GOLDSMITH, FOREIGN RELATIONS LAW: CASES AND MATERIALS 268 (4th ed. 2011).


65. Id.

66. Id. at 8. Similarly, State Department Legal Adviser, Harold Koh, testified before Congress that the Obama Administration interpreted the War Powers Resolution to apply to expansive conflicts, such as the one in Vietnam, not to smaller operations like Libya that were comparatively limited in means, objectives, risk to service members, and potential for escalation. Libya and War Powers: Hearing Before the S. Foreign Relations Comm., 112th Cong. (2011) (statement of Harold Hongju Koh, Legal Advisor, U.S. Dep’t of State), available at www.state.gov/s/l/releases/remarks/167250.htm.

67. Krass, supra note 64, at 14.

B. The War on Terrorism

The interbranch dynamic described thus far provides an essential backdrop for the focus of modern national security policy on combating terrorism. Initial counterterrorism efforts in the 1980s consisted of intelligence operations that targeted specific terrorist groups and undermined their political goals through infiltration and misinformation. 69 This approach proved well-suited to groups with specific political agendas, but it was a much less effective strategy against al-Qaeda, which emerged as a threat during the 1990s. 70 The scale of the attacks on September 11, 2001, the repeated targeting of United States’ military resources and financial center, and al-Qaeda’s declarations of holy war against Western civilization led President George W. Bush to conclude that America had been attacked, not by a fringe criminal group, but by a military enemy. 71 Not surprisingly, the war paradigm that subsequently dominated anti-terrorism policies engaged an absolutist approach to terrorism that is now perpetuated by all three branches of government.

1. Statutory Authority for Absolutist Antiterrorism Policies

The first major sign of the absolutist approach is evident, not in executive action, but in the Antiterrorism and Effective Death Penalty Act of 1996 (AEDPA), a response to recent acts of international and domestic terrorism, including the 1993 World Trade Center bombing and the Oklahoma City bombing. Title III of AEDPA aims to maximize the financial isolation of terrorist groups and is based, in part, on the finding that “foreign organizations that engage in terrorist activity are so tainted by their criminal conduct that any contribution to such an organization facilitates that conduct.” 72 Due to concerns about terrorist groups raising money in the United States “under the cloak of a humanitarian or charitable exercise,” 73 AEDPA makes it a criminal offense to provide “material support” to


70. Naftali, supra note 69.

71. Interviews show that the administration understood the United States to be at war before the last of the four planes, Flight 93, had crashed in Pennsylvania. See generally BOB WOODWARD, BUSH AT WAR 15-18 (2002).


a designated foreign terrorist organization. The current statutory language reads:

   Whoever knowingly provides material support or resources to a foreign terrorist organization, or attempts or conspires to do so, shall be fined under this title or imprisoned not more than 15 years, or both, and, if the death of any person results, shall be imprisoned for any term of years or for life. To violate this paragraph, the person must have knowledge that the organization is a designated terrorist organization . . . , that the organization has engaged or engages in terrorist activity . . . , or that the organization has engaged or engages in terrorism.⁷⁴

Some commentators believe the material support offense is a “catch-all” way to give the government room to adapt its terror prosecutions to new iterations of the threat as they emerge.⁷⁵ If so, then its inclusion was prescient in light of the attack on September 11, 2001, after which the little-utilized provision quickly became “the centerpiece of the Justice Department’s criminal war on terrorism,” despite targeting non-violent activities.⁷⁶ The statutory terms not only allow for a broad reading of “material assistance,” but also give a central and nearly unimpeachable role to the State Department’s expansive definition of terrorism.⁷⁷ The ability to bring a legal challenge to a terrorist designation is quite circumscribed under this statute, and the parties that have done so have been “almost uniformly unsuccessful.”⁷⁸

In 2010, the Supreme Court upheld the broadest possible interpretation of the “material support” offense in Holder v. Humanitarian Law Project.⁷⁹ In that case, plaintiffs were individuals and groups whose work involved teaching and advocating the use of international law and other nonviolent means to reduce conflict, advance human rights, and promote peace.⁸⁰ They raised two First Amendment chal-

---

⁷⁷. Being a non-state actor who engages in political violence is sufficient. See In re S-K, 23 I. & N. Dec. 936, 948 (BIA 2006) (“Any group that has used a weapon for any purpose other than for personal monetary gain can, under this statute, be labeled a terrorist organization.”).
⁷⁸. Wadie E. Said, The Material Support Prosecution and Foreign Policy, 86 Ind. L.J. 543, 558-60 (2011). A terrorist group does not receive notice of its designation and has thirty days to challenge it in court subsequent to publication in the Federal Register; it will only be reversed on a judicial finding that it was arbitrary, capricious, or an abuse of discretion. Id.
⁸⁰. See id. at 10.
lenges to the government’s interpretation of the material support prohibition, which criminalized their activities.\textsuperscript{81} Noting that the statute defines “material support” to include “training,” “expert advice or assistance,” “service,” and “personnel,” the Court rejected their claims.\textsuperscript{82} Writing for the majority, Chief Justice Roberts acknowledged congressional and executive findings and also independently endorsed the view that even support not related to violence “helps lend legitimacy to foreign terrorist groups—legitimacy that makes it easier for those groups to persist, to recruit members, and to raise funds—all of which facilitate more terrorist attacks.”\textsuperscript{83}

The other key piece of legislation in the U.S. legal framework for antiterrorism is the Authorization of Military Force (AUMF), passed on September 18, 2001, which gave legislative endorsement to the American military operations in Afghanistan that would begin the following month. The authorization provides:

[T]he President is authorized to use all necessary and appropriate force against those nations, organizations, or persons he determines planned, authorized, committed, or aided the terrorist attacks that occurred on September 11th, 2001, or harbored such organizations or persons, in order to prevent any future acts of international terrorism against the United States by such nations, organizations or persons.\textsuperscript{84}

This language gave the Bush Administration wide latitude to adapt traditional conventions of war conduct to an unconventional adversary. By now, the implications are familiar: framing U.S. military action against al-Qaeda as a war allowed the United States to treat members of the group as enemy combatants under international humanitarian law. Whereas criminals are entitled to due process and cannot be killed by law enforcement unless they pose an imminent threat of death or serious bodily injury, active militants in a violent conflict can be killed without warning or detained without charge while hostilities continue.\textsuperscript{85}

\textsuperscript{81} Id. at 10-11.

\textsuperscript{82} Id. at 14.

\textsuperscript{83} Holder, 561 U.S. at 30. The Chief Justice did not mention Brandenburg v. Ohio, 395 U.S. 444 (1969), which ordinarily sets an extremely high standard for government infringement on the freedom of speech. See also Hedges v. Obama, 890 F. Supp. 2d 424 (S.D.N.Y. 2012), cert. denied, 136 S. Ct. 1936 (2014) (American journalists and activists sued to enjoin the government from detaining them as enemy combatants in the event that their work would bring them into contact with designated terrorists).


2. Targeted Killing

The Obama Administration has expanded the legal principles described above to make targeted killing the foundation of U.S. antiterrorism.\(^\text{86}\) Drone strikes increased sixfold after 2009.\(^\text{87}\) Though the strikes mostly targeted Taliban fighters in Pakistan,\(^\text{88}\) the focus has recently shifted toward Yemen and (again) Iraq, with Somalia, Mali, and other weak African states on the horizon.\(^\text{89}\) The withdrawal of U.S. ground troops from Iraq and Afghanistan indicates that America’s various national security agencies are institutionalizing targeted killing policies for long-term use.\(^\text{90}\)

President Obama has attempted to provide transparency regarding the process that his Administration uses to “nominate” drone targets. A diverse group of national security officials coordinated to create a “disposition matrix” that contains information about targets and the feasibility of killing them by drone strikes or other methods.\(^\text{91}\) When the matrix suggests a good candidate for a drone strike, the group passes that information up through the President’s National Security Council.\(^\text{92}\) President Obama personally approves every name and about one-third of the total strikes.\(^\text{93}\) But not every strike targets particular individuals. Drone operations also include “signature strikes” on targets whose identities are unknown, but who intelligence shows engaging in “suspicious behavior.”\(^\text{94}\)


\(^{87}\) See Peter Bergen & Megan Braun, Drone is Obama’s Weapon of Choice, CNN (Sept. 19, 2012), http://www.cnn.com/2012/09/05/opinion/bergen-obama-drone/.

\(^{88}\) Id.


\(^{91}\) Id.


\(^{94}\) Miller, supra note 90; Greg Miller, CIA Seeks New Authority to Expand Yemen Drone Campaign, WASH. POST (Apr. 18, 2012), http://www.washingtonpost.com/
By all accounts, the drones have been extremely successful at killing individuals who the government believes are associated with terrorist organizations, and more precise targeting appears to have dramatically reduced collateral damage under the Obama Administration.\textsuperscript{95} This success has had two positive effects: (1) overwhelming domestic, public approval of the drone campaign, and (2) a significant reduction in the number of strikes.\textsuperscript{96} But despite its apparent effectiveness, the legal foundations of this counterterrorism strategy impose few, if any, requirements of transparency and accountability on the executive branch.

The Obama Administration’s position is that targeted killing is legal under three conditions: (1) the target poses an imminent threat of violent attack against the United States, (2) capture is not feasible, and (3) the operation is conducted in accordance with applicable laws of war.\textsuperscript{97} According to former Attorney General Eric Holder, if the executive concludes that these three conditions are met, then the executive’s deliberations will be enough to satisfy due process rights of suspected terrorists.\textsuperscript{98} There is no obligation to prove or defend the conclusion in a judicial proceeding.\textsuperscript{99} Accordingly, the government has not explained, for example, the risk assessments that determine whether capture is feasible,\textsuperscript{100} or how an apparently exhaustive administrative deliberation can be squared with the concept of an “imminent” threat. News reports indicate that abuses of this discretion-

\textsuperscript{world/national-security/cia-seeks-new-authority-to-expand-yemen-drone-campaign/2012/04/18/gIQAasumRT_story.html. }

\textsuperscript{95. The drone strikes have reportedly “decimated the ranks of low-level combatants,” killing as many as 2,300 suspected terrorists. Civilian casualty estimates report a drop from thirty-three percent of the total during the Bush Administration, to ten percent or less of the total casualties in 2012. Bergen & Braun, supra note 87; Drone Wars Pakistan: Analysis, Int’l Sec., http://securitydata.newamerica.net/drones/pakistan/analysis (last visited Feb. 11 2015).}

\textsuperscript{96. Micah Zenko, U.S. Public Opinion on Drone Strikes, COUNCIL ON FOREIGN REL. (Mar. 18, 2013), http://blogs.cfr.org/zenko/2013/03/18/u-s-public-opinion-on-drone-strikes/; Tracking America’s Drone War, supra note 89.}


\textsuperscript{98. Holder, supra note 97.}

\textsuperscript{99. Id. “[D]ue process does not necessarily mean judicial process.” Id.}

\textsuperscript{100. For example, the United States’ contention that it was not feasible to capture, rather than assassinate, Osama bin Laden does not seem to be verifiable. See Alan Silverleib, The Killing of bin Laden: Was It Legal?, CNN (May 6, 2011), www.cnn.com/2011/WORLD/asiapcf/05/04/bin.laden.legal/index.html.}
ary power are already taking place, such as a presumption that all military-aged men in the vicinity of known terrorist activities are enemy combatants.\footnote{101} Both the Obama Administration and Congress seem to realize that targeted killing, without effective oversight, lays a foundation for abuse of executive power—if not by the Obama Administration itself, then by its successors. As the 2012 election drew near, the Obama Administration’s officials hurried to assemble a rulebook articulating legal standards for targeted killings in the event that a new President will control those operations.\footnote{102} In addition, members of the Senate Foreign Relations Committee have tried to exercise more oversight over drone strikes. At the Committee’s insistence, each month, the Central Intelligence Agency privately screens videos of its recent drone strikes with its members (or, more commonly, their high-level aides), and it also shares a summary of the intelligence that motivated the attacks.\footnote{103}

However, as of early 2013, some members of the Committee still reported having only limited access to classified information about the strikes and their legal bases. In February, Senator Dianne Feinstein (D-CA) said that “[r]ight now it is very hard [to oversee] because it is regarded as a covert activity, so when you see something that is wrong and you ask to be able to address it, you are told no.”\footnote{104} In March, Senator Ron Wyden (D-OR) said that he was still unable to provide adequate legal oversight of drone strikes, because he still had not been given—presumably by the Obama Administration—manageable legal standards for evaluating their legitimacy.\footnote{105}

\section*{C. The Inevitable Absolutism of the Executive’s Duty to Defend}

The proposition that executives are likely to take an expansive view of their own legal authority was the primary fear that motivated the constitutional Framers to devise a government constrained by

\footnotesize
\begin{itemize}
\item 101. Becker & Shane, \textit{supra} note 93.
\end{itemize}
checks and balances. Before Harold Koh became the State Department’s Legal Adviser under President Obama, he was a law professor known for advocating a “new national security charter.”\textsuperscript{106} Years before Americans heard of al-Qaeda, Koh advocated for “attacking the institutional sources of congressional acquiescence and judicial tolerance that have contributed equally to recent executive excesses.”\textsuperscript{107} In a 1991 panel discussion, he noted, “Each of the three branches has an incentive to behave in a way which makes the system as a whole work poorly. The Executive Branch has an incentive to act or to overreach; Congress has an incentive to defer; the courts have an incentive to duck the hard cases.”\textsuperscript{108}

Similarly addressing executive overreach, Henry P. Monaghan noted the enormous public expectations that have developed around the Office of the President, pointing to the vast number of legal areas for which Americans hold presidents responsible and the speed with which they must deal with issues of law enforcement.\textsuperscript{109} Executives recognize that they, not members of Congress, will be held accountable for successful terrorist attacks. Under such pressures, Monaghan explains, “it is not surprising that ‘law’ of any kind (the Constitution included) can easily become merely one more factor to be considered, or even an obstacle to be overcome.”\textsuperscript{110}

Legal tolerance for executive dominance may also be perpetuated by a common perception that presidents and other executive officials have other, less formal incentives to exercise restraint on their own power; in the context of counterterrorism, these might include the costs and risks of military operations and the desire for professional advancement.\textsuperscript{111} President Obama portrays his own philosophical values and governing objectives as limits on his actions, and he fre-


\textsuperscript{109} Henry P. Monaghan, \textit{The Protective Power of the Presidency}, 93 Colum. L. Rev. 1, 8 (1993). Monaghan argues that although presidents cannot claim any law-making authority, the executive power implies “a general authority to protect and defend the personnel, property, and instrumentalities of the United States from harm.” \textit{Id.} at 11.

\textsuperscript{110} \textit{Id.} at 9.

quently acknowledges the need for oversight. If the executive appears to be limited in practice, then the need for formal constraints may not seem particularly pressing.

In light of the permissive attitude of Congress and the courts, the demands of the presidency, and institutional executive advantages in the realm of foreign policy, it should come as no surprise that presidents from both parties have responded to the unprecedented and difficult threat of global terrorism by invoking the most deferential legal standards possible: those that apply in times of war. But once adopted, the war paradigm, with its life-or-death stakes, is inherently absolutist. In addition, terrorists, by specifically targeting civilians, are especially likely to inspire extremely risk-averse policies that crowd out more nuanced alternatives.

Nevertheless, most academic work in the context of terrorism concludes that even if Congress can escape political responsibility for failing to address the threat, there is no compelling legal justification for its abdication of oversight in this area. Robert Bejesky characterizes the judiciary’s reluctance to arbitrate competing views of war-making authority—despite relative academic consensus on a shared-power model—as a legal void, naturally filled by the executive’s expansive understanding of its own power. Congress has likewise left its policy void with respect to the broad American objective of reducing global terrorism, as members of Congress look to the Obama Administration to tell them the legal basis for drone attacks, rather than vice versa.

III. THE COSTS OF ABSOLUTIST ANTITERRORISM POLICIES

The previous section established that current antiterrorism policies are subject to disturbingly few formal legal constraints and argued that this permissive framework inevitably leads to an absolutist view of terrorism. While only a minority of Americans may currently

113. See supra Part II.B.2.
perceive the consequences of executive dominance in this area, the following discussion offers a broad sample of the consequences of absolutism and explains why legislative checks are needed.

A. Contraction of Domestic Civil Liberties

Executive dominance in the realm of national security has encroached on individual rights to due process, counsel, and privacy against unreasonable searches. First, the Obama Administration has claimed the authority to target American citizens with drone strikes if they pose an imminent threat to the United States.\(^{116}\) When the father of such an American target, Anwar Al-Aulaqi, sued the Obama Administration on his son’s behalf for injunctive relief from assassination, a federal court held that the question of whether the government could kill Al-Aulaqi “without charge, trial, or conviction” was a “political question” and, thus, was non-justiciable.\(^{117}\) The court reasoned that deciding the issue would bleed into foreign policymaking for which the court had neither the authority nor the expertise.\(^{118}\) Al-Aulaqi was killed by drone in Yemen the following year.\(^{119}\) Responding to congressional requests to explain to Americans the circumstances under which “their government believes that it is allowed to kill them,”\(^{120}\) President Obama said that targeted killing is a tactic for U.S. military operations against al-Qaeda authorized by the AUMF, regardless of the target’s nationality.\(^{121}\)

Second, the war on terrorism has involved extensive use of “indefinite detention.” In *Hamdi v. Rumsfeld*, a plurality of the Supreme Court held that some process is still due to Americans whom the

---


117. Al-Aulaqi v. Obama, 727 F. Supp. 2d 1, 10-12, 46 (D.D.C. 2010). In 2010, the government targeted Anwar Al-Aulaqi, an American citizen, for assassination, because he had extensive ties to al-Qaeda and had allegedly played an “operational” role in the thwarted terror attack by Farouk Abdulmutallab, the “Christmas Day bomber.” *Id.* at 10.

118. *Id.* at 46. This analysis, the court said, would require resolving such matters as the precise involvement of Al-Aulaqi in al-Qaeda; whether al-Qaeda in the Arab Peninsula was linked closely enough to al-Qaeda such that hostilities against it were authorized by AUMF; whether Al-Aulaqi’s activities made him a “concrete, specific, and imminent threat to life or physical safety”; and whether the United States could employ other means to neutralize that threat. *Id.* (citation omitted).


121. Obama, *supra* note 112.
Commander-in-Chief classifies as “enemy combatants.” However, Justice O’Connor’s plurality opinion concluded that detention of enemy combatants—which, according to international law, is a “fundamental incident” to waging war—was legitimate while hostilities continued. This holding did not reconcile the Court’s rejection of indefinite detention with the perpetual nature of “hostilities” against non-state terrorist groups. The Court subsequently declined to review a D.C. Circuit decision regarding whether hostilities were continuing on the grounds that such an issue required a “political decision.” The Court in Boumediene v. Bush rejected actions by both the executive and Congress when it upheld both the executive and Congress by upholding an American detainee’s right to petition for a writ of habeas corpus. But despite the decision in Boumediene and a Guantanamo Review Task Force that has cleared some detainees for release, many detainees remain in custody without charge pursuant to an executive moratorium on the release of Yemenis.

Third, the Obama Administration has recently expanded the use of a “public safety” exception to arrestees’ rights to remain silent under Miranda v. Arizona. In 2011, the Department of Justice instructed the Federal Bureau of Investigation (FBI): “There may be exceptional cases in which, although all relevant public safety questions have been asked, agents nonetheless conclude that continued unwarned interrogation is necessary to collect valuable and timely intelligence not related to any immediate threat.” The De-

123. Id. at 519-21.
127. New York v. Quarles, 467 U.S. 649, 657 (1984) (holding Miranda requirements could be waived while there was an immediate threat to public safety, which in that case, was a missing gun).
129. Memorandum from the Fed. Bureau of Investigation on Custodial Interrogation for Public Safety and Intelligence-Gathering Purposes of Operational Terrorists Inside the United States (Oct. 21, 2010), available at http://www.nytimes.com/2011/03/25/us/25 miranda-text.html. The Department issued this internal memorandum after its request to Congress for additional guidance about applying the public safety exception went unan-
partment’s approach to interrogating Dzhokhar Tsarnaev, one of the alleged 2013 Boston Marathon bombers, suggests that this rule is the new standard for Miranda warnings. President Obama almost immediately identified the attack as an act of terrorism, and a Justice Department official said that Tsarnaev’s interrogators would invoke the public safety exception as long as needed to gain “critical intelligence.”

Finally, confidential documents recently leaked by a contractor for the National Security Agency revealed that the Agency has been operating a number of programs that allow the federal government to access and search data related to millions of Americans’ phone and Internet usage, compiled by corporations such as Verizon Wireless, Apple, Google, and Facebook. These companies are sometimes compelled by U.S. Foreign Intelligence Surveillance Court (FISA Court) orders, which were top secret until one of them was leaked in 2013, to turn over the user data that they routinely collect when a court finds that government access is justified under 50 U.S.C. § 1851. The leaked FISA Court order, for example, compelled Verizon to turn over “all call detail records or ‘telephony


132. Timothy B. Lee, Here’s Everything We Know About PRISM To Date, WASH. POST (June 12, 2013), www.washingtonpost.com/blogs/wonkblog/wp/2013/06/12/heres-everything-we-know-about-prism-to-date/.

133. FISA Courts are established pursuant to 50 U.S.C. § 1802 (2012).


135. 50 U.S.C. § 1861(a)(1) (2012) allows the FBI to apply for a court order compelling a business to produce tangibles related to an investigation of either a non-American citizen, or an American citizen in connection with international terrorism or clandestine intelligence. The order will be granted if:

there are reasonable grounds to believe that the tangible things sought are relevant to an authorized investigation . . . , such things being presumptively relevant to an authorized investigation if . . . they pertain to—(i) a foreign power or an agent of a foreign power; (ii) the activities of a suspected agent of a foreign power who is the subject of such authorized investigation; or (iii) an individual in contact with, or known to, a suspected agent of a foreign power who is the subject of such authorized investigation.

Id. § 1861(b)(2)(A). The government must also detail “minimization procedures” pursuant to § 1861(g) that are aimed at preventing the dissemination or disclosure of nonpublic information from nonconsenting individuals. See id. § 1861(b)(2)(B).
metadata’ created by Verizon for communications (i) between the United States and abroad; or (ii) wholly within the United States, including local telephone calls.\textsuperscript{136}

President Obama, defending the Agency’s search programs, said: “You can’t have 100 percent security and also then have 100 percent privacy and 0 percent inconvenience. . . . We’re going to have to make some choices as a country. What you can say is, in evaluating these programs, they make a difference to anticipate and prevent possible terrorist activity.”\textsuperscript{137} The President also observed that in light of legislative and judicial oversight of these programs, it is problematic if Americans do not trust the government’s fundamental system of checks and balances.\textsuperscript{138}

\textbf{B. Compromised Foreign Policy Goals}

Because members of violent organizations do not necessarily isolate themselves from civilian areas and may even attempt to carry out some basic governing functions, people with no intent to commit violence (e.g., family members, housekeepers, drivers, doctors, journalists) may nevertheless become linked to terrorists through social connections or economic dependence. An absolutist approach to terrorism severely disadvantages these individuals and places significant constraints on other U.S. foreign policy objectives, such as public diplomacy, refugee relief, and economic development.

For example, drone strikes in Pakistan and Yemen are inflicting collateral damage that generates hatred of the U.S. government among local populations. Although the Pakistan drone campaign is classified, most sources report that strikes in Pakistan peaked in 2010, with about 120 strikes killing between 411 and 884 civilians (including 168 to 197 children) and injuring between 1177 and 1480.\textsuperscript{139} The Pakistani government, which shares the goal of elimi-


\textsuperscript{137} Philip Ewing, NSA Memo Pushed to “Rethink” 4th Amendment, POLITICO (June 7, 2013, 2:21 PM), www.politico.com/story/2013/06/verizon-memo-4th-amendment-92416.html.

\textsuperscript{138} See id.

nating al-Qaeda and Taliban presence, must now respond to public outcry against the strikes. In April 2013, a Pakistani court said that the strikes were illegal,140 and in June, Nawaz Sharif, a staunch opponent of the drone strikes, became Pakistan’s new prime minister.141 The new Pakistani government has reportedly lodged an official protest with the American delegation to end the strikes, saying they are inspiring anti-American sentiment among the public, thus undermining the relationship between the two countries.142

Though the United States has struck Yemeni targets less frequently, some of the most highly publicized civilian casualties have occurred there, and the Bureau of Investigative Journalism reports up to 49 civilian fatalities and 144 injuries as of 2013.143 In April 2013, at a subcommittee hearing for the Senate Judiciary Committee, legislators heard testimony from Farea Al-Muslimi, a Yemeni activist who advocates for better relations with the United States.144 He reported that the strikes were undermining the Yemeni government, inspiring resentment of the United States, and giving more legitimacy to al-Qaeda groups in the region.145 “What radicals had previously failed to achieve in my village, one drone strike accomplished in an instant: there is now an intense anger and growing hatred of America,” Al-Muslimi said.146

A second policy area affected by the absolutist approach is political asylum. The legal burden to prove political persecution is high;147 however, even if refugees do not qualify for asylum status, U.S. and international law prohibits their return (“refoulement”) to countries where their life or freedom would be threatened based on their race, religion, nationality, political opinion, or membership in a particular

---

141. Shah, supra note 140.
143. Serle & Woods, supra note 139.
145. Id. at 6-7.
146. Id at 4.
147. To be granted asylum, individuals must show that they have a well-founded fear of persecution in their home country, based on their race, religion, nationality, political opinion, or membership in a particular social group. 8 U.S.C. § 1101(a)(42) (2012).
social group.148 On the other hand, current immigration laws create a complete bar to asylum if the applicant has “engaged in terrorism,” which includes providing material support to terrorist organizations.149 A terrorist organization is a group of two or more individuals, whether organized or not, which engages in any activity that is unlawful involving explosives, firearms, or any other dangerous device with intent to endanger one or more individuals or to cause substantial damage to property.150 There are no statutory exceptions based on knowledge or duress.151 Although the Board of Immigration Appeals has developed a basic mens rea requirement in its rulings, and the Department of Homeland Security can grant waivers at its discretion, recent scholarship indicates that these remedies are not reliable.152 By sending such people back to danger, the United States undermines the goals of its asylum policy and may also violate international standards for refoulement.

A third foreign policy area compromised by absolutist antiterrorism is humanitarian aid to promote economic development. In 2009, an extreme drought in the Horn of Africa led to a famine that put 3.7 million people, mostly in southern Somalia, “in crisis.”153 But an al-Qaeda-controlled Somali terrorist organization called al Shabaab, substantially prevented aid groups from delivering food154 and de-

150. § 1182(a)(3)(B).
151. Under Holder v. Humanitarian Law Project, 561 U.S. 1 (2010), this language disqualifies from asylum not only simple armed criminals, but almost anyone who has ever conducted an economic transaction with such people. See Maryellen Fullerton, Terrorism, Torture, and Refugee Protection in the United States, 29 REFUGEE SURV. Q. 4, 13 (2010) (describing the definition of terrorist organization as “two guys and a gun”).
152. This work reveals tragic stories of refugees victimized twice: by violence in their home country, and by unfair process when they seek asylum in the United States. They include individuals fleeing violent families, women engaged or married to violent men, and democracy activists charged as violent criminals by oppressive governments. See, e.g., Scott Aldworth, Terror Firma: The Unyielding Terrorism Bar in the Immigration and Nationality Act, 14 LEWIS & CLARK L. REV. 1159 (2010); Daniella Pozzo Darnell, The Scarlett Letter “T”: the Tier III Terrorist Classification’s Inconsistent and Inept Effect on Asylum Relief for Members and Supporters of Pro-Democratic Groups, 41 U. BALD. L. REV. 557 (2012); Gregory F. Lauffer, Admission Denied: In Support of a Duress Exception to the Immigration and Nationality Act’s “Material Support for Terrorism” Provision, 20 GEO. IMMIGR. L.J. 437 (2006); Courtney Schusheim, Cruel Distinctions of the I.N.A.’s Material Support Bar, 11 N.Y. CITY L. REV. 469 (2008).
154. After the provisional government took control of the country, this loose organization of Islamists fought the new regime with suicide attacks and assassinations. They also undertook to win local clan support by handing out food and money; but these efforts were undermined by forced recruitment of children, abuse of women, and kidnappings. The group pledged its allegiance to al-Qaeda in 2008. See Johnathan Masters, Backgrounders:
manded payments in exchange for access to starving populations. Relief workers, including some who worked for the U.S. government, feared criminal prosecution for material support under federal law in the event that money, food, or other resources inevitably found their way to al Shabaab. The famine worsened for two years before the U.S. Department of Treasury’s Enforcement Office agreed not to pursue “support” delivered “in good faith” to Somalia through the State Department. It also stated that food assistance was “not a focus” of agency enforcement action, but that any other person or group giving money to anyone in Somalia “should be extremely cautious.”

The Obama Administration’s slow response to the worst famine in sixty years led to a Senate Foreign Relations Committee hearing, after which Congress directed the implicated agencies to evaluate their processes. But administrative reassessment alone is unlikely to address the underlying policy dilemma: an absolute commitment to eradicating terrorists often means also punishing non-terrorists in a way that can generate political instability and anti-American sentiment, thereby actually strengthening the roots of terrorism. The United States is likely to invest in Somalia’s political environment for the foreseeable future; in the meantime, al-Qaeda affiliates continue to build strongholds wherever effective government is absent, abus-

---


157. Id.


159. Frequently Asked Q&As, supra note 158.

ing local populations in the process. The Somalia dilemma will almost certainly recur, forcing the United States to make difficult, subjective, and fact-specific choices between waging war on terrorists and promoting social stability and prosperity. But the policy nuance and flexibility that such choices require is being crowded out by the executive branch’s predictably extreme, and virtually unchecked, reliance on the use of coercive power to eliminate terrorist threats.

IV. ASSERTING A CONGRESSIONAL CHECK ON NATIONAL SECURITY POLICY

In traditional wars, threats could be reduced with the exercise of superior military force, and the model of a Commander-in-Chief simply executing a war declared by Congress was less problematic. But most political violence since World War II has taken the form of civil conflict and often involves non-governmental entities. Despite the short-term successes of the United States’ counterterrorism tactics, no one expects terrorism to disappear as a significant threat in the foreseeable future. Bruce Riedel, a counterterrorism adviser to President Obama, illustrated the futility of drone strikes as a long-term strategy: “You’ve got to mow the lawn all the time. The minute you stop mowing, the grass is going to grow back.” Because the nature of armed conflict has changed, laws based on a traditional state-to-state model of conflict—including U.S. constitutional war powers—are already becoming obsolete, thus creating space for new conventions to take their place. The rapid evolution of weapons and warfare also favors executive expertise, intelligence resources, and quick response time.

Under these conditions, Congress will have to fight for any influence it wishes to have over the nation’s foreign policy. But it has tools to do so: the Constitution provides clear authority for the legislature to set foreign policy objectives, including parameters for military actions against terrorists. Congress can do this by reforming three key statutes that currently define the legal framework of U.S. foreign policy: AUMF, the prohibition on providing material support to terrorist groups, and the Foreign Assistance Act.

163. Miller, supra note 90.
In light of how the contours of the war on terror have changed since 2001, Congress should repeal AUMF and replace it with legislation that more specifically identifies America’s enemies. The new authorization statute should codify requirements for targeted killing that provide manageable judicial standards for determining whether executive action is authorized. It should also increase transparency surrounding drone strikes by prohibiting their use in covert operations and requiring their results to be reported publicly.

After *Holder v. Humanitarian Law Project*, the prohibition on providing material support to terrorist groups should be a specific intent crime in which the perpetrator’s purpose must be to contribute to violent activities or to promote the stated goals of the terrorist organization. The prohibition should exclude groups that seek to undermine terrorism through socialization (e.g., doctors, teachers, and journalists). At the very least, the prohibition should exclude interactions that actively seek to discourage a terrorist group’s violence.

Congress should also direct the Departments of State and Homeland Security to develop rules that: (1) assess the humanitarian impact of any antiterrorism policy; and (2) set special enforcement priorities in dire humanitarian circumstances, such as natural disaster. These procedures should allow relaxed enforcement in cases of an emergency, like famine, that may call for a temporary rebalancing of priorities. Congress should also direct the State Department to compile an annual list of neutral aid agencies to be granted immunity under the material support prohibition.\(^\text{164}\)

Finally, Congress must revisit its foreign assistance framework if economic development is to be an effective complement to counterterrorism. In an unprecedented conceptual merger of strategic, ideological, and humanitarian interests after the horror of September 11, 2001, policymakers acknowledged global poverty as a threat to national security—terrorists can operate most effectively in societies with weak governments, and they can exploit the anger and desperation of poor populations.\(^\text{165}\) Yet foreign assistance is profoundly un-

---

164. Kate Mackintosh notes that groups such as Médecins Sans Frontières have long operated under a principle of neutrality that ensures neither side in a conflict will prevent their relief work. She argues that such groups must stay neutral even with respect to terrorists; otherwise, civilian populations will not get access to emergency relief and relief workers themselves will become targets of political violence. Kate Mackintosh, *Holder v. Humanitarian Law Project: Implications for Humanitarian Action: A View from Médecins Sans Frontières*, 34 SUFFOLK TRANSNAT’L L. REV. 507, 507-08 (2011).

165. As Colin Powell explained at the end of his tenure as U.S. Secretary of State, “Poverty breeds frustration and resentment, which ideological entrepreneurs can turn into support for . . . terrorism.” Colin L. Powell, *No Country Left Behind*, FOREIGN POLY (Jan. 5, 2005), http://www.foreignpolicy.com/articles/2005/01/05/no_country_left_behind; see also
dermed by development programs’ reputation for inefficiency and lack of accountability, which has engendered strong legislative mistrust. Global non-governmental organizations are seen by some as inadequate, at best, and self-serving, at worst.

To address these concerns, a congressional roundtable spent three years drafting a proposed replacement of the fifty-year-old Foreign Assistance Act. The new legislation seeks to improve efficiency and accountability through modernized reporting practices, information sharing between agencies, decentralized aid that utilizes local resources, and frequent re-evaluation of basic objectives and strategies to adjust to rapidly changing circumstances. The bill also makes specific reforms to emergency and disaster assistance programs, generally prioritizing global food security in development programs and implementing better advance planning for emergency aid to reduce inefficiency caused by haste. Although a thorough analysis of the Global Partnerships Act is well beyond the scope of this Comment, I suggest only that if legislators wish to decrease the need for military counterterrorism, they should continue these efforts to make badly-needed changes to U.S. foreign assistance.

V. CONCLUSION

Entrenched congressional and judicial deference to the executive branch in matters of national security has predictably led to an absolutist approach to combating global terrorism. This approach is implemented in a way that encroaches on American civil liberties and


166. William Easterly argues that aid organizations operate as cartels in a system where “customers” (aid recipients) have no way to provide feedback about the assistance they receive; while at the same time, no organizing entity ensures coordination and efficiency. William Easterly, The Cartel of Good Intentions, 131 FOREIGN POLY 40 (2002). Journalist Linda Polman argues that the emergency response model of many foreign aid “corporations” has incentivized the manufacture of emergencies, encouraged Band-Aid solutions, and diverted attention from more effective long-term prevention strategies. LINDA POLMAN, WAR GAMES: THE STORY OF AID AND WAR IN MODERN TIMES 10-11 (2010).

167. As aid to Africa has increased, per capita income has decreased and the number of people living on less than one dollar per day has almost doubled. Dambisa Moyo, Why Foreign Aid is Hurting Africa, WALL ST. J. (Mar. 21, 2009, 11:59 PM), online.wsj.com/article/SB123758895999200083.html. About half of food aid delivered by the World Food Program to Somalia during its recent famine was unaccounted for or misdirected, and local distribution centers operated quite differently depending on whether there were journalists present. See Katharine Houreld, Somalia Famine Aid: How Aid Went Astray, HUFFINGTON POST (Mar. 17, 2012, 10:29 AM), www.huffingtonpost.com/2012/03/17/somalia-famine-aid_n_1355348.html.


169. Id.

170. Id.
undermines the broader goal of supporting stable and prosperous societies around the world. While reasonable people may differ on the correct balance between security and liberty, the foregoing analysis demonstrates that legal restraints on the executive in this area are not particularly apparent. When the war paradigm persists with no end in sight, actions that may have once represented the outer bounds of legality can become normal standards. President Obama correctly acknowledges that the United States, as a society, should engage in a profound debate about the tradeoffs at stake in the war on terror. However, that debate should take place, not only among members of the public and the press, but also between the political branches in the adversarial system that the constitutional Framers envisioned.
THE FAILURE AND FUTURE OF
LAKE OKEECHOBEE WATER RELEASES:
A QUASI-GOVERNMENTAL SOLUTION

JACQUELYN A. THOMAS*

I. INTRODUCTION

The events that unfolded during the summer of 2013 with respect to Lake Okeechobee and the surrounding estuaries are tragic and unacceptable. The St. Lucie River and Caloosahatchee River estuaries were devastated after the Army Corps of Engineers (Army Corps) released billions of gallons of polluted fresh water from Lake Okeechobee into the estuaries during a particularly rainy season.¹ The outdated Herbert Hoover Dike surrounding Lake Okeechobee can only withstand so much water pressure, and with water levels approaching the maximum level of safety, the Army Corps chose to release lake water into the estuaries to the east and west rather than risk a breach of the dike, which would flood cities and farmland to the south.²

During and after these events, local citizens, nonprofit environmental organizations, and local and state politicians were vocal in their collective opposition to, and disapproval of, the Army Corps'...
choices regarding Lake Okeechobee water releases. The Army Corps maintains plenary jurisdiction over the dike and the regulation of water releases under the Rivers and Harbors Act of 1899 and the Clean Water Act (CWA), and there are allegations that the agency has failed in its mission to properly regulate and maintain the infrastructure that keeps this delicate ecosystem in a constant state of flux. Many people called for reforms, and in February 2014, the Florida Senate sent a letter to Congress asking that it transfer authority over water releases from the Army Corps to the Florida Department of Environmental Protection (FDEP).

A full assessment of the failed regulation of Lake Okeechobee and the surrounding estuaries is beyond the scope of this Note. Rather, this Note focuses on a narrower aspect of the problem: the Army Corps’ plenary jurisdiction over the Herbert Hoover Dike and the regulation of water releases into the estuaries. It discusses the advantages and disadvantages of retaining jurisdiction in the Army Corps versus transferring jurisdiction to the FDEP, and it reaches the following conclusion: neither is best suited to manage this problem. Therefore, this Note proposes the transfer of jurisdictional power from the Army Corps to a new quasi-governmental commission composed of various interested parties—public and private—at the federal, state, and local level. This commission would make decisions about the regulation and maintenance of Lake Okeechobee and the Herbert Hoover Dike. Ideally, it would also oversee the implementation of short- and long-term restoration plans—both of which are already being discussed—to stop or significantly reduce the water released into the estuaries. There are several reasons why this solution is preferable to either the Army Corps maintaining jurisdiction or the FDEP obtaining jurisdiction. As will be discussed in greater depth below, a quasi-governmental organization would: (1) provide management flexibility and efficiency, (2) represent and incorporate the interests of various parties, (3) utilize the institutional knowledge and expertise of those who are closer to the resource, and (4) have the

---

3. See infra Part III.
5. See infra notes 25-31 and accompanying text.
ability to obtain private funding in a time of public funding reductions and constrained budgets.

Part II briefly summarizes the history of Lake Okeechobee and the events that unfolded in 2013 during and after the water releases that overwhelmed the estuaries. Part III discusses the history of the Army Corps’ jurisdiction over navigable waterways, including its expansion under the CWA. Then, Part IV analyzes arguments for and against reserving the Army Corps’ jurisdiction over Lake Okeechobee water releases and the Herbert Hoover Dike. It also argues for a cooperative federalism regime in which a quasi-governmental organization is formed, which, as will be explained, is the best solution for balancing the various interests of all parties involved. Part V highlights what such a transfer of power would look like and how a quasi-governmental organization could be formed. Part VI concludes.

II. THE HISTORY AND CURRENT REGIME SURROUNDING THE LAKE OKEECHOBEE BASIN

A. How Did We Get Here?

The regulation of the Lake Okeechobee Basin is complex, and the problem has existed for decades.7 At 730 square miles, Lake Okeechobee is the largest lake in the southeastern United States, though it is extremely shallow, with an average lake-wide depth of nine feet.8 Lake Okeechobee naturally receives water from the north, “from a watershed . . . that includes the Upper Kissimmee Chain of Lakes, the Kissimmee River,” and other smaller lakes, creeks, and drainage basins.9 The Eleventh Circuit recently explained the history and current state of the lake, which is worth quoting at length:

Historically, the lake had an ill-defined southern shoreline because during rainy seasons it overflowed, spilling a wide, shallow sheet of water overland to the Florida Bay. “But progress came and took its toll, and in the name of flood control, they made their plans and they drained the land.”

In the 1930s the Herbert Hoover Dike was built along the southern shore of Lake Okeechobee. It was intended to control flooding but failed during the hurricanes of 1947 and 1948. Con-

gress then authorized the Central and Southern Florida Flood Project; as part of it the Army Corps of Engineers expanded the Hoover Dike and built pump stations including S–2, S–3, and S–4. Under the modern version of that project, nearly all water flow in South Florida is controlled by a complex system of gates, dikes, canals, and pump stations.

The area south of Lake Okeechobee’s shoreline was designated the Everglades Agricultural Area. The Corps dug canals there to collect rainwater and runoff from the sugar cane fields and the surrounding industrial and residential areas. Not surprisingly, those canals contain a loathsome concoction of chemical contaminants including nitrogen, phosphorous, and un-ionized ammonia. The water in the canals is full of suspended and dissolved solids and has a low oxygen content.

Those polluted canals connect to Lake Okeechobee, which is now virtually surrounded by the Hoover Dike.\(^\text{10}\)

The above passage describes how pollution makes its way into the lake. This is an important and controversial problem in itself and has been the subject of multiple lawsuits.\(^\text{11}\) However, this Note focuses on what happens next in the regulation of this intricate water system.

When the water level of Lake Okeechobee reaches its limit, the Army Corps releases the polluted fresh water into the delicate St. Lucie River and Caloosahatchee River estuaries.\(^\text{12}\) This is nothing new; the Army Corps has been overseeing water releases for many years. The difference is that during the summer of 2013, heavy rain caused the water level to rise so quickly that the Corps was forced to open the proverbial floodgates and inundate the estuaries with polluted fresh water.\(^\text{13}\) The estuaries are composed of “brackish water with higher salinity levels [to] support a delicate ecosystem. The fresh water lowered the salinity levels and oysters, sea grasses and other wildlife began dying.”\(^\text{14}\) Toxic algae bloomed, and water conditions were so poor in August that signs were posted to warn people to...
stay out of the water. One only needs to see the photographs of brown water pouring out of the lake and into the rivers; the massive, toxic brown and green algae blooms overtaking the naturally vibrant blue water; and dead flora and fauna (including manatees) to comprehend the seriousness of this problem. The releases have had negative effects on the region’s economy, as well.

However, it is important to note that water releases from Lake Okeechobee are not the sole cause of this delicate ecosystem’s decline. “Rather it has been a combination of factors that have resulted in what some scientists have referred to as ‘the perfect storm.’” Comprehensive regulation of the lake and its natural and artificial tributaries is necessary to solve this problem. Water releases into the surrounding estuaries, however, have a direct negative impact on the health of these ecosystems, and the releases need to be addressed as part of a complete restoration plan.

In any event, residents are angry, frustrated, and powerless, and the overlapping participation of several entities can be confusing. Is Congress, the Army Corps, the U.S. Environmental Protection Agency (EPA), the Florida Legislature, the state’s Governor, FDEP, or the South Florida Water Management District (SFWMD) to blame? Depending on whom you ask, the fault allocation among the various entities is likely to change. At the very least, the federal government is not solely to blame, even if Florida Governor Rick Scott currently disagrees. For example, when Charlie Crist was governor, the state began constructing a water reservoir for use as an alternative to releasing water into the estuaries, but construction was later abandoned after spending millions of dollars on it. Also, environmentalists point out that Governor Scott and the Florida Legislature have cut funding to the SFWMD in recent years and that “inexperienced managers” have been appointed to the governing board. And Governor Scott only turned his attention to Lake Okeechobee after the

15. Id.
16. Devastating Photos of Florida Pollution Will Fill You with Rage, HUFFINGTON POST (Oct. 3, 2013, 8:52 AM), http://www.huffingtonpost.com/2013/10/02/lake-okeechobee-pollution_n_4031154.html. The photographs come from Congressman Patrick Murphy’s official Facebook page. Last summer, he asked local residents to send him photos of the devastation, which he then compiled into an album of over one hundred photos entitled, “Show Congress the Crisis of Our Waterways.” See id.
17. See SELECT COMMITTEE FINAL REPORT, supra note 6, at 10; Reid, supra note 7.
18. SELECT COMMITTEE FINAL REPORT, supra note 6, at 1.
19. Id.
20. Reid, supra note 4.
events of last summer. Nevertheless, the fact remains that the Army Corps maintains plenary jurisdiction over the Herbert Hoover Dike and water releases into the estuaries. Hopefully what happened last summer will finally act as the catalyst for reform; not piecemeal, ad hoc fixes that deal only with immediate issues, but real reform about the manner in which this extremely important, delicate, and complex water system is regulated.

B. The Senate Select Committee on Indian River Lagoon and Lake Okeechobee Basin

Due to the growing concern over environmental damage from continued water releases and questionable decision-making by the Army Corps, the Florida Senate Select Committee on Indian River Lagoon and Lake Okeechobee Basin (IRLLOB) was formed on July 10, 2013. The Select Committee’s purpose was to review water management policies in the Basin, assess the environmental impacts of water releases, identify options for improvement at the state and federal level, and “[d]evelop recommendations for improved water management.” The Select Committee released a report on November 8, 2013, in which it “recommends amending the operational jurisdiction of the Army Corps of Engineers to give the State of Florida, specifically the Department of Environmental Protection, authority over regulatory releases when the risk of dike failure is less than 10 percent.” In addition, when the risk of dike failure is greater than ten percent, the report discusses procedures for providing the Army Corps with 24-hours’ notice before the State of Florida decides whether to maintain control or to temporarily cede control to the Army Corps. As a result of the Select Committee’s report, State Sena-

23. See id.

24. The South Florida Water Management District published its Final Adaptive Protocols for Lake Okeechobee Operations in 2010, which were developed in cooperation with the Army Corps and the FDEP. The introduction summarizes the evolution of decision-making regarding Lake Okeechobee water releases from the early 1990s to present and highlights the relationship between the SFWMD and the Army Corps. It is true that the two entities have worked together to undertake studies of the issue, but in the end, the SFWMD (or any other entity) is limited to making recommendations, requests, and suggestions to the Army Corps. The day-to-day decision-making remains solely within the federal agency’s jurisdiction and control. See S. FLA. WATER MGMT. DIST., FINAL ADAPTIVE PROTOCOLS FOR LAKE OKEECHOBEE OPERATIONS iii-v (2010) [hereinafter SFWMD FINAL ADAPTIVE PROTOCOLS], available at http://www.sfwmd.gov/portal/page/portal/xrepository/sfwmd_repository_pdf/ap_lo_final_20100916.pdf.


26. Id.

27. SELECT COMMITTEE FINAL REPORT, supra note 6, at 21.

28. Id.
tor Joe Negron, chair of the Select Committee, wrote a letter to Congress asking that it remove the Army Corps’ sole jurisdiction over the lake.\textsuperscript{29} Senator Negron stated that the Army Corps “has [recently] demonstrated a willingness to be more proactive and coordinate with the South Florida Water Management District to manage lake levels.”\textsuperscript{30} He also noted that the Army Corps uses its plenary power to make decisions “even when those actions conflict with the state water managers’ better judgment” and thus asked Congress to “rebalance this delegation of responsibility and authority.”\textsuperscript{31} As of this writing, Congress has yet to respond.

The solution proposed in this Note is thus similar to the Select Committee’s proposal in that it seeks to divest the Army Corps of plenary jurisdiction, but it is also different in several fundamental respects that will be addressed later. Furthermore, while the Select Committee’s proposal seems simple and straightforward, in actuality, removing jurisdiction from the Army Corps and placing it in a different entity is serious and complex, and it would alter the current balance of federal and state power. The origins of this power are discussed in the following Part.

III. JURISDICTION OVER NAVIGABLE WATERWAYS

In addition to the duties the Army Corps performs for the Army in times of conflict, its civil engineering jurisdiction has evolved, contracted, and expanded over the last two centuries. Currently, the Army Corps maintains jurisdiction over certain important aspects of water resources management, including the authorization and regulation of all structures built in, on, or across navigable waters of the United States pursuant to section 10 of the Rivers and Harbors Act of 1899.\textsuperscript{32} The Army Corps also maintains control over all dredging and filling of navigable waters under section 404 of the CWA.\textsuperscript{33} Historically, the Army Corps’ jurisdiction has covered transportation (e.g., bridges and roads), civil projects (e.g., lighthouses, public buildings, and monuments), and surveying (e.g., much of the West and the Great Lakes).\textsuperscript{34} However, this Note limits its focus to the evolution


\textsuperscript{30} Id. at 1-2.

\textsuperscript{31} Id. at 2.


and expansion of the Army Corps’ control over “navigable waters of the United States.” In order to understand the current state of affairs regarding the regulation of the Lake Okeechobee Basin, it is important to review precisely how the Army Corps came into power in the first place.

A. The Case Law Defining Navigability

The U.S. Constitution indirectly mentions water only once; Article III, Section 2 states that “all Cases of admiralty and maritime jurisdiction” are within federal judicial control. What this means is that admiralty cases must be heard in federal court; state courts do not have jurisdiction to hear these disputes. Admiralty jurisdiction is important “because of the seminal role it played in defining ‘navigability’ in early U.S. law, and the subsequent adoption of some of the same principles in other definitions of navigable waters.” Essentially, the definition of “navigable waters,” developed by the courts in the nineteenth century to define the scope of federal admiralty jurisdiction, was exported to other areas of federal law when Congress sought to expand the scope of its power over the country’s waters.

In England, a special court existed that heard only admiralty disputes at the time the U.S. Constitution was drafted, and the English court’s jurisdiction was limited to waters affected by the ebb and flow of the tide. In the early years of the Republic, American courts essentially adopted the English definition and narrowly construed federal admiralty jurisdiction to encompass only coastal and tidal waters. This limitation was abandoned in Waring v. Clark, which acted as a catalyst for the expansion of the definition of navigability in the United States. In The Propeller Genesee Chief v. Fitzhugh, the Supreme Court demonstrated the effects of Waring by adopting a new, broader definition of the term navigability. This change was due in part to the realization and acceptance that the geography of the United States was far different than that of England; for example, the Court noted that the Mississippi River is navigable well beyond the location at which the tide ceases to have any effect. It would be “purely artificial and arbitrary as well as unjust” to draw a line across the Mississippi River, as commerce took place on the river

35. See Clean Water Act § 404; Rivers and Harbors Act § 10.
38. Id.
39. Id. at 283-84 (collecting cases).
40. 46 U.S. (5 How.) 441 (1847).
41. 53 U.S. (12 How.) 443 (1851).
42. Id. at 456-57.
above and below this location.\textsuperscript{43} The Court’s new, broader definition made jurisdiction “depend upon the navigable character of the water, and not upon the ebb and flow of the tide. If the water was navigable it was deemed to be public; and if public, was regarded as within the legitimate scope of the [federal] admiralty jurisdiction.”\textsuperscript{44} Thus, the important factor for the Court was whether the water body could maintain commerce, not whether the water body was tidal in nature. This was an important and necessary first step in the federal government’s desire to expand its regulatory power over U.S. waters.

In a separate doctrinal field nearly three decades earlier, the Supreme Court clarified the ambiguity between commerce and navigation in \textit{Gibbons v. Ogden}.\textsuperscript{45} There, the Court held that limiting accessibility to a state’s navigable waters gives rise to a Commerce Clause violation, as this unconstitutionally infringes upon Congress’s power to regulate commerce.\textsuperscript{46} In response to the challenge that navigation is not a commercial transaction, the Court explained that so long as navigation has the capability to affect interstate commerce, it can be regulated under Congress’s Commerce Clause power.\textsuperscript{47} Thus, this decision, combined with the Court’s later opinion in \textit{The Propeller Genesee Chief}, opened the door for the Supreme Court to address the scope of the navigability doctrine—as previously developed in admiralty law—in the context of the Commerce Clause.\textsuperscript{48} In \textit{The Daniel Ball}, the Supreme Court put a name to its new test: navigability in fact.\textsuperscript{49} More specifically, the Court stated that “[t]hose rivers must be regarded as public navigable waters in law which are navigable in fact.”\textsuperscript{50} Further, “they constitute navigable waters of the United States . . . when they form in their ordinary condition by themselves, or by uniting with other waters, a continued highway over which commerce is or may be carried on with other States . . . .”\textsuperscript{51} Thus, the test has two elements that must be established before the federal government can claim jurisdiction over the water body: (1) it must be in its ordinary condition, and (2) it must form a continued highway for commerce.\textsuperscript{52} These elements have been construed quite broadly in later cases, however.\textsuperscript{53}

\begin{itemize}
\item \textsuperscript{43} \textit{Id.} at 457.
\item \textsuperscript{44} \textit{Id.}
\item \textsuperscript{45} 22 U.S. (9 Wheat.) 1 (1824).
\item \textsuperscript{46} \textit{See id.} at 89-91.
\item \textsuperscript{47} \textit{See id.}
\item \textsuperscript{48} \textit{ADLER, CRAIG & HALL, supra note 37, at 289.}
\item \textsuperscript{49} 77 U.S. (10 Wall.) 557, 563 (1870).
\item \textsuperscript{50} \textit{Id.}
\item \textsuperscript{51} \textit{Id.}
\item \textsuperscript{52} \textit{ADLER, CRAIG & HALL, supra note 37, at 291.}
\item \textsuperscript{53} \textit{See id.} at 291-92 (citing examples of the Court’s broad construction of these elements).
\end{itemize}
In addition to the expansion of Congress’s jurisdiction over the physical location of navigable waters, the Court in United States v. Appalachian Electric Power articulated the full scope of Congress’s Commerce Clause powers over several ancillary issues of navigability.\(^{54}\) Encompassed within the concept of commerce for navigability purposes is more than just the regulation of boats on the water; commerce also includes flood control, watershed management, and the creation of electricity sources.\(^{55}\) As a result of this evolution in the case law, much of the ambiguity surrounding the limits of federal jurisdiction over waters of the United States was resolved, and the federal government thus gained the potential for control over most of this country’s waters.\(^{56}\)

B. The Clean Water Act

Using its articulated power, Congress enacted a statute in 1948 that attempted to address the problems of water pollution throughout the United States.\(^{57}\) The CWA in its current form is the result of amendments to the original statute in 1972.\(^{58}\) The amendments over the years progressively gave more regulatory power to the federal government, and the 1972 amendments solidified federal regulatory control over certain aspects of water management and pollution control of navigable waters.\(^{59}\) The term “navigable waters” is defined in the CWA as “the waters of the United States, including the territorial seas.”\(^{60}\) This term has been litigated many times, and the Supreme Court most recently articulated its definition in Rapanos v. United

---

54. 311 U.S. 377, 426 (1940).
55. Id. In addition, the Supreme Court held that “[t]o appraise the evidence on navigability on the natural condition only of the waterway is erroneous. Its availability for navigation must also be considered.” Id. at 407 (emphasis added). The Court explained that some artificial improvement to a waterway to make it navigable in fact (and the question of what constitutes too much improvement) is “a matter of degree.” Id. The improvements need not be completed or even begun before jurisdiction attaches. Id. at 408. Also, it is important to note the Court’s statement that “[o]nce found to be navigable, a waterway remains so.” Id.
56. This author says the “potential” for control because, for reasons that may seem obvious, the federal government must first affirmatively determine whether a particular water body is “navigable” before it may assert jurisdiction over it. Once a water body is deemed navigable, certain statutory regimes take effect, and the federal government has expressly preempted state jurisdiction (e.g., CWA NPDES permitting). For other types of regulation concerning a navigable water that are not expressly preempted by Congress, there may still be dormant preclusion under the Supreme Court’s holding in Gibbons v. Ogden, 22 U.S. (9 Wheat.) 1 (1824).
59. ADLER, CRAIG & HALL, supra note 37, at 557-58.
60. 33 U.S.C. § 1362(7).
States. In that case, the Supreme Court narrowed the EPA’s broad definition of “waters of the United States” in regards to wetlands adjacent to a traditional navigable water; however, the extent of the limitation is not certain. This is because the main opinion only consisted of a plurality of four Justices, led by Justice Scalia, who articulated a narrow definition of “waters,” while Justice Kennedy’s concurrence explicated a different, broader, and more fact-specific definition of the term. Justice Kennedy’s test is known as the “significant nexus” test, because it requires “the existence of a significant nexus between the wetlands in question and navigable waters in the traditional sense. . . . [If the wetlands . . . significantly affect the chemical, physical, and biological integrity of other covered waters more readily understood as ‘navigable,’] then that water body may also be regulated under the CWA. As a result of the Supreme Court’s failure to reach a majority, states were free to adopt either Justice Scalia’s narrower definition or Justice Kennedy’s broader definition. A circuit split has since emerged, with most circuits either adopting Justice Kennedy’s test or holding that either test applies.

The two most well-known provisions of the CWA are sections 402 and 404. Section 402 regulates discharges of pollutants from point sources, which excludes agricultural and other nonpoint source runoff. Essentially, if an entity wishes to discharge pollutants into a water of the United States, it must obtain a National Pollutant Discharge Elimination System (NPDES) permit from the EPA or from a state agency that, pursuant to a cooperative federalism provision in the statute, has been authorized and delegated to run the program. Such a provision requires that the state agency demonstrate to the EPA that it is qualified to implement the program before the EPA will cede jurisdiction.

62. Id. at 732. Specifically, Justice Scalia’s definition limits “waters” to include “only relatively permanent, standing or flowing bodies of water.” Id. Importantly, this definition limited the Army Corps’ section 404 jurisdiction over wetlands that are only saturated intermittently, not permanently. See id.
63. Id. at 779-80 (Kennedy, J., concurring).
64. Id.
67. Id.
68. Id.
to manage the NPDES permitting program within the state.\textsuperscript{69} In addition, Congress left control over nonpoint source pollution (e.g., runoff from the land and atmospheric deposition of pollutants) to the states to manage as each sees fit, so long as minimum federal water quality standards are met.\textsuperscript{70} In Florida, the FDEP develops these water quality standards and implements them with the aid of the five water management districts.\textsuperscript{71} The second provision of the CWA, section 404, encompasses “the discharge of dredged or fill material into the navigable waters” of the United States, which requires a permit from the Army Corps.\textsuperscript{72} In sum, “[f]rom a federalism perspective, Congress designed the Clean Water Act to reflect a balance between national uniformity in some respects and individual state needs in others.”\textsuperscript{73} Confusion regarding overlapping jurisdictional authority has not been entirely eradicated, however, as the resource in question is not always easy to conceptualize, quantify, and manage; that is, water is a fluid resource. The regulation of the Lake Okeechobee Basin provides a good example of this ambiguity and jurisdictional overlap.

In the early twentieth century, the Army Corps built a complex system of “spillways, locks, pump stations, culverts, canals, reser-

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{69} \textit{National Pollutant Discharge Elimination System (NPDES), Fla. Dep’t of Envtl. Prot.}, http://www.dep.state.fl.us/water/stormwater/npdes/index.htm (last updated Dec. 19, 2014).
\item \textsuperscript{70} See 33 U.S.C. §§ 1313, 1329, 1362(14); see also \textsc{Adler, Craig & Hall}, supra note 37, at 561.
\item \textsuperscript{72} 33 U.S.C. § 1344(a). Similar to section 402, the CWA provides states the opportunity to take over section 404 permitting—to an extent. Unlike the NPDES permitting program, if a state wishes to administer its own section 404 permitting regime, its jurisdiction only extends to those waters deemed “non-navigable” under federal law. See id. § 1344(g); see also State or Tribal Assumption of the Section 404 Permit Program, U.S. EPA, http://water.epa.gov/type/wetlands/outreach/fact23.cfm (last updated July 1, 2014). From this author’s reading of the statutory provision and EPA’s explanation on its website, this limitation maintains the status quo—that is, the Army Corps retains jurisdiction over all section 404 permitting for navigable waters of the United States, and the states may administer dredge-and-fill permitting for non-navigable waters over which the federal government never had jurisdiction anyway. It appears that what this “cooperative federalism” provision does is simply allow states to utilize and incorporate the federal section 404 regime as its own, which is why authorization is required from the EPA. See id. However, it is important to note that states do retain some power over section 404 permitting pursuant to section 401 of the CWA, which requires state certification of compliance with federal and state water quality standards before a federal permit may be issued. See 33 U.S.C. § 1341; see also PUD No. 1 of Jefferson Cnty. v. Wash. Dep’t of Ecology, 511 U.S. 700 (1994).
\item \textsuperscript{73} \textsc{Adler, Craig & Hall}, supra note 37, at 561.
\end{enumerate}
\end{footnotesize}
voirs, and water conservation areas” in South Florida.\textsuperscript{74} Interestingly, the operational control of this system has since been transferred from the Army Corps to the SFWMD.\textsuperscript{75} Among other things, the network of canals and pump stations artificially diverts agricultural, industrial, and residential runoff away from the agricultural lands to the south and carries the runoff to Lake Okeechobee, which pollutes the water that is periodically released into the estuaries when the lake’s water level reaches its limit.\textsuperscript{76} It is important to note that because the polluted canal water being pumped into the lake comes from nonpoint source pollutants, there is no NPDES permit requirement by the EPA or FDEP, and there is no Army Corps permit requirement, as the pollutants are not considered dredge or fill material. The FDEP and SFWMD regulate nonpoint source pollution under the provision of the CWA that requires states to create and implement water quality-based standards, which allows for the inclusion of nonpoint source pollution when deciding how to enforce them.\textsuperscript{77} The FDEP thus regulates pollutants from all sources, and for those water bodies that are “impaired,” the agency must determine the “total maximum daily load” (TMDL) for each pollutant and allocate the allowable daily amount among all of the water body’s polluters.\textsuperscript{78} The main pollutant


75. Operational Planning, SFWMD, http://www.sfwmd.gov/portal/page/portal/xweb%20-%20release%20%20operational%20planning (last visited Jan. 18, 2015) (“With approximately 2,100 miles of canals and 2,000 miles of levees/berms, 70 pump stations and more than 600 water control structures and 625 project culverts, the District actively operates and maintains the water management system to protect regional water supplies and provide flood control for 8.1 million people—plus the environment, agriculture, businesses and visitors—in South Florida.”); see also Friends of the Everglades v. S. Fla. Water Mgmt. Dist., 570 F.3d 1210, 1214 (11th Cir. 2009) (“Those polluted canals connect to Lake Okeechobee, which is now virtually surrounded by the Hoover Dike. The S–2, S–3, and S–4 pump stations are built into the dike and pump water from the lower levels in the canals outside the dike into the higher lake water. . . . At full capacity, the pumps within the S–2, S–3, and S–4 stations can each move 900 cubic feet of water per second—more than 400,000 gallons per minute. The South Florida Water Management District operates the pumping stations.”).

76. See supra text accompanying notes 10-12.

77. See STORMWATER/NONPOINT SOURCE MGMT. SECTION, supra note 71, at 14-16.

78. See id.; see also 33 U.S.C. § 1313(d) (2012); Everglades: Lake Okeechobee, FLA. DEP’T OF ENVTL. PROT., http://www.dep.state.fl.us/everglades/lakeo.htm (last updated May 17, 2013). After several years of litigation involving the EPA, FDEP, and environmental nonprofits, the FDEP finalized updated numeric nutrient criteria (i.e., TMDLs) of several nutrients for the various water bodies throughout the state, including nitrogen and phosphorus in Lake Okeechobee and the estuaries, in 2013. The updated criteria went into effect on September 17, 2014, when the EPA formally withdrew its competing numeric nutrient criteria. See EPA, NUMERIC NUTRIENT CRITERIA FOR THE STATE OF FLORIDA: WITHDRAWING THE FEDERAL ACTIONS (2014), available at http://water.epa.gov/lawsregs/rulestregs/upload/Numeric-Nutrient-Criteria-for-the-State-of-Florida-Withdrawing-the-Federal-Actions-Factsheet.pdf; see also Development of Numeric Nutrient Criteria for Florida’s Waters, FLA. DEP’T OF ENVTL. PROT., http://www.dep.state.fl.us/water/wqssp/nutrients/
of concern in the Lake Okeechobee Basin is phosphorous, which continues to be discharged into the lake at a rate exceeding that which is necessary to maintain a healthy ecosystem.\footnote{Everglades: Lake Okeechobee, supra note 78.} Only time will tell whether the FDEP’s newly updated TMDL for phosphorus has an appreciable effect on lowering the level of this pollutant in the Lake Okeechobee Basin.

In 2002, environmental nonprofits filed suit against the SFWMD in an attempt to limit the water management district’s authority to freely pump polluted canal water originating from nonpoint sources. Specifically, in \textit{Friends of the Everglades v. South Florida Water Management District},\footnote{Id. at 1216 (citing 33 U.S.C. §§ 1311, 1342(a)(1), 1362(12)).} the parties argued over the definition of “discharge of a pollutant” in relation to NPDES permitting requirements. A NPDES permit is required for the “discharge of any pollutant,” and “discharge” is defined as “any addition of any pollutant to navigable waters from any point source.”\footnote{Id. at 1213.} The plaintiffs argued that “the transfer of a pollutant from one navigable body of water to another [constituted] a ‘discharge of a pollutant’ within the meaning of the Clean Water Act.”\footnote{See id. at 1216.} In other words, the plaintiffs believed that the artificial pumping stations used to pump polluted water from the canals into Lake Okeechobee constituted a separate “point source,” such that the SFWMD should be required to obtain a NPDES permit before discharging pollutants into a navigable water—here, the lake itself.\footnote{Id. at 1217-18 (discussing decisions of other circuits).} The Eleventh Circuit noted that other circuits had previously agreed with similar arguments and ruled in favor of requiring NPDES permits.\footnote{Id. at 1218.} And it appears that the Court agreed in theory with this line of decisions; however, the EPA had recently promulgated a new regulation that took the opposite view.\footnote{Id. at 1218-19 (quoting NPDES Water Transfers Rule, 73 Fed. Reg. 33,697 (June 13, 2008) (codified at 40 C.F.R. § 122.3(i))).} The EPA’s “water transfers” rule exempts water transfers from NPDES permitting, “‘defin[ing] water transfers as an activity that conveys or connects waters of the United States without subjecting the transferred water to intervening industrial, municipal, or commercial use.’”\footnote{Id. at 1218-19 (quoting NPDES Water Transfers Rule, 73 Fed. Reg. 33,697 (June 13, 2008) (codified at 40 C.F.R. § 122.3(i))).} The Eleventh Circuit undertook a \textit{Chevron} analysis of this new rule, and it determined that the definitions within the CWA were sufficiently
ambiguous to move on to step two.\textsuperscript{87} Thus, so long as the agency’s interpretation was reasonable, which the Court found it to be, the Court was bound to give effect to the rule.\textsuperscript{88}

This case demonstrates the complexity of the overlapping jurisdiction of the Lake Okeechobee Basin, with different entities having control over different aspects of the system’s resources. While the Army Corps has control over the water releases into the estuaries, the FDEP and the SFWMD have control over the levels of phosphorus and other pollutants that run off into the canals, whose waters ultimately end up in the lake. Also, the Army Corps transferred operational authority over the complex network of physical infrastructure to the SFWMD, and the water management district provides the Army Corps with much of the data that it uses to make decisions.\textsuperscript{89} Thus, these groups currently work in coordination with one another, but only to a certain extent. For example, the Army Corps did not consult with the SFWMD, FDEP, or any other entity before choosing to release the deluge of water that caused last summer’s devastation to the estuaries.\textsuperscript{88} The following Part addresses this federal, state, and local jurisdictional overlap and argues that jurisdiction should be transferred to a new, quasi-governmental entity.

\section*{IV. Federalism and the Quasi-Governmental Organization}

In its basic form, federalism is “the distribution of power in an organization (as a government) between a central authority and the constituent units.”\textsuperscript{90} In the United States, federalism describes the balance of power and authority between the federal government and the states, based on concepts and boundaries articulated in the U.S. Constitution.\textsuperscript{92} In the context of environmental law generally, and water law more precisely, federalism plays an important role in de-

\textsuperscript{87} Id. at 1222-27.
\textsuperscript{88} Id. at 1219-20, 1227-28.
\textsuperscript{89} See Operational Planning, supra note 75 (“Operational Planning includes the review and development of operating protocols and plans, coordination with District groups and the U.S. Army Corps of Engineers regarding Operating Manual review and development. Operational Planning uses the current state of the system (Lake Okeechobee levels and groundwater levels) as input to the South Florida Water Management Model (SFWM). A suite of graphics are produced for the decision makers to analyze. The application involves the investigation of potential areas of flexibility in the South Florida Water Management District water control system operational guidelines and federal regulation water level schedules. The objective is to improve the water resource evaluation tools available to water managers to develop a comprehensive set of operational performance measures and to provide a forum for public input to operations.”).
\textsuperscript{90} See Select Committee Final Report, supra note 6, at 10, 21.
fining the boundaries of federal and state control. In the context of this Note, the federal government expressly preempted Florida from regulating in some areas, while in other areas, Congress paved the way for a cooperative federalism regime in which Florida was allowed to take over regulatory authority after receiving approval from the EPA or the Army Corps. Still in other areas, Congress affirmatively required Florida to take the lead role. But what about quasi-governmental organizations—where do they fit into this picture? Before addressing this question in depth, the following section briefly describes the quasi-governmental organization.

A. Public-Private Partnership: The Quasi-Governmental Organization

Simply put, quasi-governmental organizations are “hybrids”—that is, “entities that combine characteristics of public- and private-sector organizations.” More specifically, they “combine authority from more than one level of government, whether as a formal or informal part of their structure or governance process, and also include private and public actors within the governance process.” Hybrid organizations purportedly combine the “best of both worlds,” so to speak, by providing the efficiency and resources of the private market and the accountability of the public domain. In addition, hybrid organizations are each unique; they exist on a continuum of organizational power vis-à-vis government regulatory oversight and control, public-private monetary funding, and composition of the organization’s board members. In 2003, there existed more than fifty quasi-governmental organizations at the federal level and several hundreds at the state and local level. These entities oversee many important interests and perform essential functions for the government and society, some more prominent than others (e.g., the Port Authority of New York and New Jersey versus the management of railroads), and as of 2003 these organizations’ collective liability was calculated at over two trillion dollars. In the realm of environmental regula-

93. Id. at 1143-44.
96. Id. Of course, there are many criticisms of quasi-governmental organizations, as well, including the potential for a lack of public accountability. The advantages and limitations of adopting this entity will be discussed in more detail infra Part IV.B.
97. Osofsky & Wiseman, supra note 95, at 1-2, 9.
98. KOPPELL, supra note 94, at 2.
99. Id.
100. Id.
tion, “cooperative environmental decision-making processes exist at all levels of government—irrespective of community and/or ecosystem size—and in various private and quasi-governmental organization guises. Some of these processes are largely defined by geographic setting, others by resource use and/or protection, and still others by statutory authorization or requirement.”

One might reasonably ask: Where does the authority for hybrid organizations come from? Are these organizations even constitutional? The answer is ambiguous, much in the same way that the answer regarding the constitutionality of the modern administrative state remains unsettled. Article I, Section 8, clause 18 of the U.S. Constitution is known as the “Necessary and Proper Clause.” It states that “Congress shall have the power . . . [t]o make all Laws which shall be necessary and proper for carrying into Execution the foregoing Powers, and all other Powers vested by this Constitution in the Government of the United States, or in any Department or Officer thereof.”

Since the country’s early history, this clause was interpreted to justify the creation of administrative agencies that aid Congress and the President in the execution of Congress’s statutory regimes and policy goals. This malleable provision has also been the justification for Congress’s experimentation with hybrid entities, which have expanded as the administrative state has expanded and the need for flexibility has increased.

There are proponents and critics of hybrid organizations, and each side offers logical and valid opinions. Likewise, there are groups on both sides of the debate concerning whether the Army Corps should retain or cede jurisdiction over water releases from Lake Okeechobee. Currently, the Lake Okeechobee Basin is already operating in a system that is largely cooperative. The goal of this Note is to shift the balance of power among the various actors so as to reach a more efficient and effective result. This shift is two-fold: it calls for an incremental increase in the amount of non-federal regulatory authority, plus a shift from state government to quasi-government. The remainder of this Part discusses several reasons for and against this shift, and it describes the quasi-governmental organization’s potential to leverage a greater amount of expertise, access to private capital, and superior flexibility and responsiveness. It also builds a case that responds to two critiques—democratic accountability and the possibility of capture.

103. KOPPELL, supra note 94, at 5.
104. See id. at 5-6.
B. Point/Counterpoint: Army Corps Versus FDEP Versus Quasi-Governmental Organization

There are many reasons why those who feel negative impacts locally may be opposed to a distant and disinterested federal agency (or, at least, that is how some perceive federal agencies) making important decisions for which it does not suffer the consequences. On the other hand, there are valid reasons for maintaining the status quo, which include institutional knowledge gained from decades of regulation, the monetary resources of the federal government, and the clarity in having one agency make this decision.

First, there is some credence to the argument that the Army Corps has superior institutional knowledge. The agency has been regulating water releases and maintaining the Herbert Hoover Dike for decades, and it knows how to do the job. The agency knows how to calculate and interpret the data on weather patterns, water levels, and other factors affecting the water body, and it has technology in place to monitor lake conditions on a constant basis. However, as mentioned above, the SFWMD also provides important data, both for itself and for the Army Corps. The Army Corps has also ceded control over the complex web of physical infrastructure that it built throughout South Florida to the SFWMD, and the latter is successfully managing “approximately 2,100 miles of canals and 2,000 miles of levees/berms, 70 pump stations and more than 600 water control structures and 625 project culverts” for more than eight million people. Thus, it appears that the SFWMD has institutional knowledge and capacity similar to the Army Corps in this area. As Robert Percival points out, part of the reason that the federal government began preempting the states in the realm of environmental law was because the states had failed to implement their own laws and policies, even after prodding from Congress. While this was certainly true in many respects fifty years ago, state and local governments have since modernized and expanded, and many have gained expertise and competence in regulating environmental issues themselves. This is especially true for water law; around forty states implement the federal NPDES permitting program, and all fifty states oversee the cre-

106. See Operational Planning, supra note 75.
107. Id.
108. Percival, supra note 92, at 1147.
109. The rise of state expertise and competence in regulating aspects of the environment is partially, if not mostly, due to cooperative federalism provisions in which Congress encouraged the states to take action by offering monetary incentives, allowed them to take over federal statutory schemes after receiving the necessary approval, or required states to implement programs at the state and local level. See id. at 1173-75.
ation and implementation of state water quality standards.\textsuperscript{110} There
can be no doubt that the federal government has the power to regu-
late navigable waters under the Commerce Clause and the federal
navigational servitude,\textsuperscript{111} but it does not necessarily follow that the
federal government \textit{must or should} act on that power if important fed-
eral interests are not at stake and other sub-federal entities have suf-

cient capacity to regulate.

The hybrid organization that this Note proposes would enhance,
rather than hinder, the expertise required to properly manage the
dike and water releases from the lake. It would require members of
the Army Corps and the SFWMD who possess this important
knowledge and skill set to be part of the governing board, thus ren-
dering transaction costs in transferring the information negligibl\textsuperscript{e.\textsuperscript{112}}

Also, outside experts from industry and non-governmental organiza-
tions—such as engineers, hydrologists, and biologists—would be part
of the commission, which only adds to the level and diversity of ex-
pertise. These individuals would be able to provide and analyze data
that the Army Corps currently does not in order to reach a more
complete understanding of the dike’s capabilities and the ability of
the estuaries to withstand such releases. When dealing with a waters-
shed as complicated as this one, it is important to have input and
analysis from many sectors. The Florida Senate Select Committee
criticized the Army Corps for recently failing to consider the eco-

cological effects of its actions,\textsuperscript{113} and the hybrid organization would address
this valid concern.

The next argument for maintaining the status quo stems from
the federal government’s vast monetary resources. The idea here is that
the federal government has more funds available to it than state gov-
ernments, and federal funding is usually necessary to fully imple-
ment an environmental program.\textsuperscript{114} There is evidence that state

\textsuperscript{110} See \textit{id.} at 1174-75.

\textsuperscript{111} The federal navigation servitude is always lurking beneath the depths of every

navigable water—or water with the potential for navigability—of the United States. It

“effectively insulates the federal government from otherwise legitimate takings claims

where federal projects or operation of federal regulatory authority in navigable waters

impair private property rights in those waters.” \textit{ADLER, CRAIG \\& HALL, supra} note 37, at

299-300.

\textsuperscript{112} Indeed, the SFWMD and the Army Corps have already worked extensively toget-

her in conducting studies of the issue, and the SFWMD has published its own final adaptive

protocols concerning Lake Okeechobee operations, which are “intended to provide opera-
tional guidance to the” water management district’s staff and Governing Board. \textit{See

SFWMD FINAL ADAPTIVE PROTOCOLS, supra} note 24, at iii-v.

\textsuperscript{113} \textit{See SENATE COMMITTEE FINAL REPORT, supra} note 6, at 10 (“[M]inimizing the risk

of dike failure to ensure public safety is the primary concern when lake levels increase.

This concern trumps all other considerations, including environmental harm.”); \textit{see also id.}

at 12 (“The Corps has been criticized over the past decade . . . for its disregard for the envi-

ronmental damage some water resource projects have caused.”).

\textsuperscript{114} \textit{Percival, supra} note 92, at 1175.
budget cuts, combined with a reduction in federal financial assistance and a failure of state governments to replace these lost funds with state funds, leads to state regulatory programs that fail to meet federally required environmental standards. When considering the Lake Okeechobee Basin, it is true that the Army Corps has the potential for far more funding than the State of Florida could provide. However, the fact remains that Congress must first appropriate funds to the agency, and the President must then allocate funds to this project in particular. In the past, the federal government has failed to allocate the funds necessary to fix the Herbert Hoover Dike and build other infrastructure that would help reduce the flow of water into the estuaries, such as water reservoirs. Indeed, Congress’s failure to provide funding for additional water reservoirs is one of the main criticisms the State of Florida has against Congress and the Army Corps. The Florida Legislature has pledged to split the financial burden of construction with the federal government, and Governor Scott claims that Congress has not been living up to its end of the bargain. While the federal government’s knee-jerk reaction might be to reject any proposal that limits its power, Congress and the Army Corps should soon realize that this reduces the burden on an overworked agency with a limited budget.

115. Id.

116. See Senate Committee Final Report, supra note 6, at 12 (summarizing a 2012 study that “found that the Corps faces an unsustainable situation driven by budgetary considerations” and that if current funding levels from Congress remain, “degraded performance” will occur).

117. See id. at 15; Alvarez, supra note 1.

118. In its 2014 budget, the Florida Legislature appropriated 231 million dollars in funding for the Indian River Lagoon and Lake Okeechobee Basin projects—including funding for additional water reservoirs and a series of southward bridges to allow some water to flow naturally into the Everglades again—by not only approving the Senate Select Committee’s monetary recommendations, but exceeding them. See Press Release, Fla. Senate, Senator Negron Announces Full Funding of Lagoon Recommendations (Apr. 28, 2014), available at http://www.flsenate.gov/Media/PressReleases/Show/1785.

119. See Alvarez, supra note 1.

120. The President’s budget for the Army Corps for fiscal year 2014 was released on March 3, 2014. In total, the budget includes roughly 4.5 billion dollars in gross discretionary spending for civil works projects. However, the President only allocated 75 million dollars to Herbert Hoover Dike seepage control and 66 million dollars to the more-general South Florida Everglades Restoration Program. See News Release No. 14-002, U.S. Army Corps of Eng’rs, U.S. Army Corps of Engineers Releases Work Plans for Fiscal Year 2014 Civil Works Appropriations (Mar. 4, 2014), available at http://www.usace.army.mil/Media/NewsReleases/NewsReleaseArticleView/tabid/231/Article/475460/us-army-corps-of-engineers-releases-work-plans-for-fiscal-year-2014-civil-works.aspx. Based on the Army Corps’ own figures, the cost of repairing and stabilizing the dike is around two billion dollars, and the agency has already spent 400 million dollars on repairs since 2007. Select Committee Final Report, supra note 6, at 8-9. Considering that the repairs are supposed to be completed by 2018, it can be inferred that appropriating 75 million dollars to control dike seepage for fiscal year 2014 falls far short of the annual amount of money required to work expeditiously to secure the dike. See id. In addition, the Select Committee’s report highlights many recent Army
The current state of affairs, therefore, does not necessarily tip the scales in favor of reserving jurisdiction in the Army Corps, but it also does not necessarily imply that the state government is preferable. Jonathan Koppell notes that this is where quasi-governmental organizations have frequently found their functional justification.121 Hybrid organizations may be funded by both public and private funds, or they may even be fully privatized.122 This Note’s proposal would call for a combination of public and private funding to best meet the large fiscal demand necessitated by dike repairs and other operations. This would ensure that the federal and state governments would still have incentives to oversee that the hybrid organization is properly managing its funds, but it also allows the inclusion of desperately needed private capital to make up the continuous public funding shortfall.123 It also provides a buffer against changing political tides, as future politicians who come into power may decide that the health of Lake Okeechobee and the St. Lucie and Caloosahatchee estuaries are low on their list of priorities.

Also, there is a third argument that having one agency—the Army Corps—control the dike and the decision over water releases actually supports simplicity, clarity, and efficiency through structural norms and accountability.124 In a similar fashion, this would be the argument made in support of the transfer of jurisdiction to the FDEP rather than a quasi-governmental organization, as jurisdiction would still remain in a single governmental body within an established hierarchical structure. There may be fears that a hybrid organization would lead to inefficiency and deadlock based on its composition by persons with varying and opposing interests.125 However, the reality of the situation is that jurisdiction in the Lake Okeechobee Basin is already split among various actors. The entire watershed is managed in a cooperative manner between federal, state, and local govern-

Corps civil works failures and seeming lack of concern for environmental damage caused by its authorization of projects. See id. at 12.

121. KOPPELL, supra note 94, at 6 (“Budget constraints and rules have always been a significant factor in the explanation for the growth of American quasi-government.”).


123. KOPPELL, supra note 94, at 3.

124. See KOSAR, supra note 122, at 31 (describing “[a] unified executive structure, coupled with hierarchical lines of authority and accountability”). But see id. at 32 (“Those favoring the public law approach to management . . . believe that the democratic value of political accountability should take precedence over the managerial value of maximizing efficiency and outcomes.”).

ments, and the federal government has already ceded control to the FDEP (which then delegated authority to the SFWMD) to manage the extremely important infrastructure that the Army Corps itself had built. Thus, the argument that jurisdiction over the Herbert Hoover Dike and the water releases into the estuaries must rest within the sole discretion of the Army Corps is without much merit. Nevertheless, this is also precisely the reason why this author does not agree with Senator Negron’s proposal to transfer jurisdiction to the FDEP. While the FDEP’s expert knowledge about the environmental issues occurring within the state are to be desired, the fact remains that this would allow one entity to control water releases without any input from other entities—essentially what is happening now with the Army Corps, just at the state level. This is contrary to the ultimate goal of managing the Basin cooperatively, effectively, and efficiently.

Proponents of hybrid organizations argue that including various interested parties in the decision-making process from all levels of government and the private sector is preferable and actually more efficient and meaningful than overlapping, ambiguous jurisdictional boundaries.126 While it may seem counterintuitive, several sectors can be involved while also increasing flexibility and responsiveness.127 This is because this type of organization requires cooperation between horizontal actors who all have the ability to influence outcomes.128 Also, the hybrid organization’s foundational procedures would be structured in such a way that deadlock would be made very difficult.129 The fears of those who criticize hybrid organizations out of concern that clashing opinions will lead to deadlock are thus likely overstated.

It is also important to think about the precedent that such a transfer of jurisdiction might set. Depending on one’s views, this could be a good or bad thing. For those who prefer power to be centralized in the federal government, the transfer of control over the Herbert Hooker Dike and Lake Okeechobee’s water releases might be seen as the beginning of a slippery slope. Their argument would be that by allowing a quasi-governmental entity to gain control over actions that have historically been within the sole purview of the federal government, this would open the door for other states that are also

126. Osofsky & Wiseman, supra note 95, at 10-12.
127. Kosar, supra note 122, at 31-32.
128. See id. at 10.
129. See id. at 11 (“Recent scholarship on new governance informs [the authors’] assessment of how hybrid structures can be designed to include stakeholders effectively and appropriately. New governance views regulation not as solely top-down, public control by state and federal agencies with central authority, but rather as an ongoing and ever-changing relationship—often one of negotiation and compromise—between agencies, regulated entities, and other stakeholders.”).
dissatisfied with the Army Corps’ regulatory decision-making. To be sure, the Army Corps does not have a perfect track record; the agency “has been criticized over the past decade for underperforming or failing civil works projects” and has been accused of displaying a lack of concern for negative environmental impacts caused by its projects.130 Also, the slippery slope argument only matters if, as a substantive matter, reserving jurisdiction in the Army Corps is the desirable alternative. For others, the hybrid organization would be a welcome solution to a federal agency that has arguably failed in its duty to properly manage this water system,131 as well as a Congress that appears apathetic and repeatedly fails to appropriate sufficient funds to build additional infrastructure and thus prevent further environmental degradation.132

There also exists a notable policy reason for ceding control to a sub-federal governmental entity, which is based on the argument that state and local governments should be able to manage their own resources as they see fit.133 The idea here is that the federal government is detached and removed from the reality of various situations and differences on the ground, and a federal agency’s decisions are sometimes contrary to those that state or local governments think are best.134 Of course, there are good reasons for having federal control in many instances, including the “guarantee [of] a minimum level of environmental protection to citizens regardless of their place of residence” and the prevention of states engaging in competition that

130. SELECT COMMITTEE FINAL REPORT, supra note 6, at 11-12.

131. The Army Corps recently completed a five-year project in which the most vulnerable portion of the dike was repaired, but there are many more repairs that need to be done. See Alvarez, supra note 1; Reid, supra note 4. Furthermore, the Army Corps has invested 400 million dollars since 2007 to make improvements to the dike, see SELECT COMMITTEE FINAL REPORT, supra note 6, at 8-9, and that it will continue to work diligently with “a team of engineers, hydrologists, geologists, scientists, contract and real estate specialists, budget analysts, and many others . . . to ensure the very best rehabilitation strategies are applied to the dike today and in the future.” U.S. ARMY CORPS OF ENG’RS, JACKSONVILLE DIST., HERBERT HOOVER DIKE REHABILITATION: PROJECT UPDATE (Spring 2013), available at http://www.saj.usace.army.mil/Portals/44/docs/FactSheets/HHD_FS_Rehab_Spring2013_508.pdf. Nevertheless, the dike remains one of the most vulnerable in the country, and this increased risk of failure drives the Army Corps’ decisions to release polluted water into the estuaries. See Alvarez, supra note 1; Reid, supra note 4. State and local officials emphasize that the area is one tropical storm away from a potentially catastrophic disaster. Alvarez, supra note 1.

132. See supra note 120 and accompanying text.

133. Percival, supra note 92, at 1144 (“State and local governments argue that federal regulations infringe on their autonomy and sovereignty, and that they impose costly unfunded mandates states can ill afford. Even though Congress has taken care to ensure that federal environmental law rarely preempts state standards, states argue that they should be given more freedom and flexibility to develop environmental standards tailored to local circumstances.”).

134. Id.
creates a race to the bottom. This would not be the case here, however, as the problem does not affect interstate pollution and involves no competition among states; rather, this involves the transfer of control of a single water body to some type of sub-federal entity (here, a hybrid organization) that is capable of managing it.

C. Criticisms of the Hybrid Model

Every idea has its drawbacks and weak points, and the quasi-governmental organization is no exception. Two criticisms of this model are worth mentioning: democratic accountability and the possibility of capture.

Fundamentally, democratic accountability is based on the idea that elected government officials can be held accountable for their actions by the citizenry. When a politician makes a mistake or a decision with which her constituents disagree, she may be voted out of office. Concerns over democratic accountability regarding the traditional administrative state have been discussed in great depth, but the issue with quasi-governmental organizations is that they go one step further than traditional administrative agencies; they are even more attenuated and protected from the potentially negative effects of the electorate. Critics of the hybrid form ask: What if the committee makes a mistake or an unpopular decision? To whom would the committee be held accountable? These are not easy questions, and the answer remains ambiguous. However, the beauty of the quasi-governmental organization is in its flexibility—depending on its structure, a hybrid organization can be subject to more or less regulatory oversight and more or less accountability. In the case of the Lake Okeechobee Basin, this author believes that the committee should maintain a fair amount of independence in its management of the dike and water releases, in the same way that an agency’s actions are entitled to deference and are only overturned if arbitrary, capricious, or otherwise not in accordance with law. However, there still must be some governmental oversight, especially because the committee will be partially composed of governmental actors and public

---

135. Id. at 1171-72.
136. KOPFELL, supra note 94, at 3.
137. KOSAR, supra note 122, at 3-5.
138. See id.; see also KOPFELL, supra note 94, at 3-4.
139. KOSAR, supra note 122, at 2 (describing the relationship of hybrid organizations to Congress or the Executive Branch as “a descending scale from closest to the most distant”).
140. 5 U.S.C. § 706(2)(A) (2012). There is also the issue of sovereign immunity and whether hybrid organizations should be entitled to it. Because the organization would still receive substantial public funding and be partially compiled of representatives from different levels of government, this author believes that the committee should be entitled to the protection of sovereign immunity.
funds will be utilized. The extent and type of oversight would need to be fleshed out in more detail, but at the least, an annual review of the quasi-governmental organization's actions and allocation of funds should be required.

The second critique to consider is that of the theory of the firm, and more specifically, the possibility of capture.\textsuperscript{141} Opponents may argue that because the hybrid organization would attract private funding and be partially composed of individuals from the private sector—both for profit and not for profit—the committee could become captured by perverse, self-serving private interests.\textsuperscript{142} It could become profit-driven rather than goal-driven. However, the hybrid organization would be created as a nonprofit entity, thus largely eliminating this threat.\textsuperscript{143} There are several additional buffers that would help prevent this from occurring, as well, including the existence of public-sector board members, maintenance of some level of regulatory oversight by the federal government, and the ability to remove individuals from the governing board.

In sum, there are valid reasons for and against ceding jurisdiction to the FDEP or a quasi-governmental entity. The above discussion analyzes these issues and provides support for the transfer of jurisdiction to a hybrid organization rather than preserving the status quo in the Army Corps or transferring jurisdiction to the FDEP. In the next Part, this Note fleshes out how this organization could be composed.

V. THE QUASI-GOVERNMENTAL COMMITTEE OF THE LAKE OKEECHOBEE BASIN

A quasi-governmental committee composed of various interested parties at the federal, state, and local levels—public and private—should be created to oversee the regulation of Lake Okeechobee water releases and the related infrastructure, including the Herbert Hoover Dike. This entity would represent the collective decisions of a group of people, each representing different interests, who are able to analyze and discuss a situation before taking action. The interests of one party would weigh against the interests of a different party, and ideally, these parties would discuss the various alternatives and reach a consensus. It may very well be the case that the committee, if presented with a similar situation to the one the Army Corps faced last summer, would make the same decision in light of public health and

\textsuperscript{141} See Osofsky & Wiseman, supra note 95, at 63.
\textsuperscript{142} See id. (“Public-private hybrids must always be alert to concerns about industry capture.”).
\textsuperscript{143} Of course, this does not entirely address the threat of capture in quasi-governmental organizations that are for-profit, but that problem is not at issue here.
safety concerns.\textsuperscript{144} Yet, rather than have different entities continue to point fingers at one another for any mistakes or shortcomings, this entity would take responsibility as a collective group.

Quasi-governmental organizations are vast and varied, and each entity’s structure is unique.\textsuperscript{145} Nevertheless, rough classifications have been created that sufficiently encompass many of the organizations at the federal level.\textsuperscript{146} The type of organization that would best suit the needs of Lake Okeechobee and the estuaries is similar to an entity classified as a nonprofit “organization independent of, but dependent upon, an agency of the federal government.”\textsuperscript{147} These organizations are mission-oriented\textsuperscript{148} rather than profit-driven, and they exist in large part because of the flexibility and necessity of funding from both public and private sources.\textsuperscript{149} Nonprofits can maintain surplus funds and invest them into the organization’s future plans; they simply cannot be used to benefit the governing board or owners.\textsuperscript{150} This would work well for the Herbert Hoover Dike; if the organization were to become profit-driven, the committee would likely appear too privatized and its governing board self-serving. The question of whether Army Corps and/or SFWMD employees would continue to work on the physical infrastructure or whether the committee could hire workers from private construction and engineering firms to complete some of the work would need to be decided, but this is not a sticking point.

The proposed committee would be composed of eleven individuals, and it would include two representatives each from the Army Corps, FDEP, and SFWMD, and one representative each from local government, an environmental nonprofit, the agricultural industry, the fishing/tourism industry, and a private engineering firm.\textsuperscript{151} Each gov-

\textsuperscript{144} See SFWMD Final Adaptive Protocols, supra note 24, at iv (The 2008 Lake Okeechobee Regulation Schedule “made . . . clear that the issue of public health and safety regarding the integrity of the Herbert Hoover Dike was the dominant factor in the decision making process to select a preferred alternative regulation schedule.”).

\textsuperscript{145} KOPPELL, supra note 94, at 2.

\textsuperscript{146} KOSAR, supra note 122, at 2-3.

\textsuperscript{147} Id. at 16.

\textsuperscript{148} Koppell notes that there are four types of principal preferences, which can be classified as (1) positive, mission; (2) negative, mission; (3) positive, non-mission; and (4) negative, non-mission. KOPPELL, supra note 94, at 71-75. He finds that negative, non-mission preferences are most likely to be satisfied, while positive, mission preferences are least likely. Id. at 75. This makes intuitive sense, though, as it is easier not to do something inconsequential than it is to actively do something important, as the latter requires much more work. The Lake Okeechobee Basin Committee would be fulfilling a positive, mission-based preference; it should be evident that the task will not be easy.

\textsuperscript{149} KOSAR, supra note 122, at 16-17.

\textsuperscript{150} Definition of ‘Not For Profit,’ INVESTOPEDIA, http://www.investopedia.com/terms/n/not-for-profit.asp (last visited Jan. 18, 2015).

\textsuperscript{151} It is important to note that this is only a proposal, and the composition of the governing committee could obviously be tweaked.
ernmental entity would choose its own representatives. Once chosen, the public-sector representatives would choose the private-sector representatives. This would help to preserve government accountability, which distinguishes this committee from a traditional nonprofit organization independent of, but dependent upon, the government.152 Also, the enabling statute would provide more detailed procedures for selection, term limits, and removal.

The committee could meet weekly, bi-weekly, or monthly, which should not be difficult given the widespread use of teleconferencing. In addition, the group should create a set of rules and regulations beyond those specified in the enabling statute to aid in self-governance and coordination of activities. This could be done by utilizing notice and comment rulemaking in order to ensure that the public has a voice.153 In times of disagreement or stalemate, there should be a process by which a tie is broken, and, potentially, a process for vetoing a decision. In times of urgency—for example, when the committee must make an immediate decision about whether to release water into the estuaries due to rising lake water levels—there should be a plan in place to convene the group immediately (requiring a certain number of participating members to form a quorum) and to quickly make decisions based on scenarios and outcomes already calculated and anticipated by the group.154 There should also be a method by which the governing board may seek assistance from the federal or state government in times of emergency, such as when a tropical storm or hurricane hits.

Ideally, if this experiment proves to be successful, the concept of hybrid organizations in this area of regulation may take hold elsewhere. Other states may seek to petition Congress for transfers of jurisdiction to newly created entities that could manage and oversee each particular state’s expensive, contentious, and important water resources and infrastructure. Further down the road, Congress may feel comfortable creating a provision of the CWA that outlines a procedure for the creation of hybrid organizations to oversee “state waters of significant concern,” much in the same way section 402 of the

152. See KOSAR, supra note 122, at 16-17. One criticism of these organizations is that they need more regulatory oversight and accountability regarding management of funds. Id. As discussed above, the federal government would be able to review all of the committee’s spending annually, which should eliminate this concern.


154. The Army Corps already makes decisions now “based on a set of quantitative performance measures of ecosystem and water supply conditions that have a strong foundation in population ecology, regional environmental science, and water resources engineering.” SFWMD FINAL ADAPTIVE PROTOCOLS, supra note 24, at iv-v. Thus, it should not prove difficult to transfer and utilize these measures, or to create updated measures, once the quasi-governmental committee is formed.
CWA outlines the procedure for transferring jurisdiction over the NPDES permitting program from the EPA to the states.

VI. CONCLUSION

This Note’s goal is to propose the best—not the perfect—regulatory solution to this very real and urgent problem. The Lake Okeechobee Basin and the St. Lucie and Caloosahatchee estuaries require action now, and Congress continues to withhold the funds necessary to properly repair the Herbert Hoover Dike and build additional water reservoirs. The Army Corps is also delayed in its schedule of repairs. Florida has allocated 230 million dollars to Basin projects in 2014, but it simply does not have enough money in its coffers to cover the entire cost—and why should it, when the state does not have control over the infrastructure and plays no role in the decision-making process regarding water releases? The time has come, then, to think critically and innovatively about the best solution moving forward, which is to create a quasi-governmental committee to manage the Herbert Hoover Dike and water releases from Lake Okeechobee into the St. Lucie and Caloosahatchee estuaries. This proposal seeks to address the critical needs of the estuaries while balancing jurisdiction among various interested parties with limited governmental oversight. It would allow the committee to seek funding from private capital and would aid in management flexibility and efficiency.